



Att'y. Docket No.: 3124-Z

#15
Attachment

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of

Richard J. Feldmann

Serial No. 09/866,925

Group Art Unit 1645

Filed: May 30, 2001

Examiner John S. Brusca

RECEIVED

JUN 11 2003

TECH CENTER 1600/2900

For: ALGORITHMIC DETERMINATION OF FLANKING DNA SEQUENCES
THAT CONTROL THE EXPRESSION OF SETS OF GENES IN
PROKARYOTIC, ARCHEA AND EUKARYOTIC GENOMES

DECLARATION UNDER 37 C.F.R. 1.132

Hon. Commissioner of Patents & Trademarks
Washington, D. C. 20231

Sir:

I, James V. Oberthaler, whose address is 7701 Woodmont Avenue,
Apt. 406, Bethesda MD, 20814, declare as follows:

1. I received a BS degree in Mathematics from the College of
the Holy Cross. I also did graduate work in Mathematics at the
University of Maryland and the American University, completing all
requirements for an MA in Mathematics except the dissertation. I
hold a Masters degree in Business Administration from Southern
Illinois University at Edwardsville, IL. My work experience includes
over 40 years of software design and engineering and management of
these activities. My relevant experience is described briefly in the
Curriculum Vitae that accompanies this letter. I currently serve as
Principal Research Analyst, under contract to the National Cancer

Institute Center for Bioinformatics.

2. I have read the patent specification for application Serial No. 09/866,925 as filed in the United States Patent & Trademark Office on May 30, 2001 for "ALGORITHMIC DETERMINATION OF FLANKING DNA SEQUENCES THAT CONTROL THE EXPRESSION OF SETS OF GENES IN PROKARYOTIC, ARCHEA AND EUKARYOTIC GENOMES." I read the amended claims submitted October 30, 2002, and I have read the amended claims submitted with the amendment accompanying this Declaration. I have read the official communication from the U.S. Patent & Trademark Office dated January 8, 2003. I have considered all of the claims.

3. I wish first to direct my comments to claims 20 - 27 which, I have been advised, are the broadest claims in the application. I have been also advised that claims 28 - 37 are all dependent from claim 20 and, hence, include the limitations of claim 20.

Claims 20 - 27 are as follows:

20. A method of identifying DNA sequences that control the expression of different collections of genes in a genome comprising detecting, by computer, one or more pairs of non-adjacent DNA sequences to which are bound one RNA molecule comprising of two RNA sequences.

21. A method of identifying DNA sequences that control the expression of different collections of genes in a genome comprising detecting, by computer, changes in connectron behavior in the genome as a function of changes in the sequence of the genome.

22. A method of modifying, by computer, the expression of different gene collections in a genome, comprising detecting changes in connectron behavior that results in changes in the level of connectron control sequences caused by an exogenous stimulus.

23. A method of detecting, by computer, where and when new genes have been integrated into a host genome comprising detecting the operable link between the newly introduced gene and the existing connectron behavior in said host genome.

24. A method of detecting, by computer, the expression effect of different gene collections in a given host genome, comprising detecting the transacting behavior of connectrons between the chromosomes thereof.

25. A method of modifying a given genome comprising modifying, by computer, the connectron organization therein.

26. A method of detecting, by computer, connectron control and target sequences in a given genome comprising:

determining the base composition of said genome,
determining one or more sites of control sequence organization,
and/or
determining one or more sites of target application.

27. A method of determining, by computer, the response of a cell in any tissue to changes in the cell's environment and/or genetic composition comprising providing a complete genomic DNA sequence for the organism and determining the effect of changes in connectrons due to application of a given exogenous stimulus to the genome.

4. It should be noted that all of the claims, including the dependent claims, are directed to a tetradic relationship between two specific adjacent RNA single-stranded sequences (called C1 and C2 for control sequence 1 and control sequence 2), which interact with two distant double-stranded DNA sequences (called T1 and T2 for target sequence 1 and target sequence 2). Claim 20 refers to a computer algorithm for identifying DNA sequences that control the expression of different collections of genes in a genome by identifying one or more pairs of non-adjacent DNA sequences to which are bound one RNA molecule containing two RNA sequences.

5. Referring to pages 5 and 6 of the Examiner's action and in reference to subparagraphs a) through h), I wish to state the following:

In subparagraph a), the Examiner first states that in order to practice the claimed invention, one skilled in the art must identify and use a connectron to predict the relation of gene expression. Keeping in mind that the claims under consideration are directed to computer mediated methods of analysis of connectron sequences, I disagree with the Examiner's conclusion for the following reasons:

(1) Claim 20 asserts that this patent application provides a mechanism (i.e., a computer algorithm) whereby one skilled in the art (i.e., a journeyman, molecular biologist, bioinformatician, or computer programmer who understands the storage format, content and use of readily available bioinformatics resources) can write software, following the algorithm, that will analyze the DNA sequence of an organism to identify DNA sequences (called C1, C2, T1 and T2 in the description of the algorithm) meeting specific criteria set forth in the description. And, further, that the identified sequences behave in such a way that when the control sequence containing C1 and C2 is transcribed into RNA, the RNA will seek out and bind to the target sequence (C1 binding to T1 and C2

binding to T2) to achieve the effect that the entire DNA sequence beginning with T1 and ending with T2 is shielded from transcription

(2) The software, implemented following the algorithm and set to work on standard, readily available genome sequences, will identify the collection of DNA and RNA sequences making up connectrons as defined in the patent application without any need for understanding by or help from the computer programmer.

Further, in subparagraph b), the Examiner states that the description provides guidance to identify connectron symmetries in genomic sequences, and I agree. However, the Examiner also contends that the description does not provide detailed guidance to use identified connectron symmetries to predict an effect on gene expression, and with respect to this contingent I disagree for same reasons stated in paragraphs (1) and (2) above.

In subparagraph c), the Examiner contends that the description provides working examples of identification of connectron symmetries in genomic sequences, and I agree. See pages 37 and 38 of the specification for a listing of the examples, and pages 39 - 188 for detailed expositions of sample uses of the algorithm. The Examiner

further contends that the description does not provide working examples of using identified connectron symmetries to predict effects on gene expression.

I disagree. On the contrary, this is exactly what the examples provide. As explained in the introduction and in the definitions provided, (particularly, the definitions of Possible Connectron and Hierarchy of Connectron Action) each connectron control sequence C1-C2 will, when transcribed into RNA, seek out and bind to its target sequence T1-T2, thereby shielding the DNA between T1 and T2 from transcription. Since the shielded DNA sequence will not be transcribed, any genes in the span between T1 and T2 will not be expressed as proteins for as long as the C1-C2 sequence remains bound to T1-T2. Similarly, any additional C1-C2 sequences in the span between T1 and T2 will also remain inactive during this time period, and therefore the inhibitory effect they otherwise would have exerted on their target sequences will be suppressed during this time period. Granted that the full, cascading sequence of transcription/expression and sequestration that would result from each of the examples discussed is not presented, the principles are given that would enable anyone who understands the mechanism, as explained in the application, to follow the effects as deeply as he or she desires.

In subparagraph d), the Examiner states that the nature of the invention, gene expression control, is complex. I agree for the reasons stated in the preceding paragraph and for the even more fundamental reason that the molecular-biological processes of even the simplest cell are very complex. Life is very complex: a fully formed organism with incredibly complicated biological activities develops from a single cell and lives a full lifetime by interacting in countless ways with its external environment. It would often be impossible to enumerate to completion all the effects that even a single connectron turning on or off would cause. Some cells would cycle through millions of possible states before repeating one. The only reasonable presentation of these effects is to give the principles that apply, and this has been done clearly and completely in the application.

In subparagraph e), the Examiner asserts that the prior art does not show connectrons; and for the purposes of this Declaration, I am assuming that the connectrons have the definition given above. I agree with the Examiner's contention. Mattick does not show connectrons as defined in the instant specification. It is my understanding and belief that the connectron invention disclosed in

the present application was made by the inventor, Richard J. Feldmann.

In subparagraph f), the Examiner contends that the skill of the art of gene expression is high, and I agree.

In subparagraph g), the Examiner contends that the predictability of the relationship of connectron symmetries and gene expression is unknown in the prior art, and I agree. Although I do not present myself as an expert in genetics or molecular biology, it is clear from the nature of the publications appearing at the time this letter is written that laboratory researchers are only now beginning to encounter, case by case, the effects **disclosed** by Mr. Feldmann's fully-formed connectron invention. To the best of my knowledge, these investigators are discovering individual applications of the invention, but no one except Mr. Feldmann has yet **disclosed** the overarching theory and its implications.

In subparagraph h), the Examiner contends that the claims are broad in that they are drawn to identification of connectron symmetries whose relationship to gene expression is not established. While I am not cognizant of the legal terms or definitions for the

breadth of claims, my understanding of the breadth of method claim 20, for example, is that it requires detecting a DNA sequence that controls the expression of different collections of genes in a genome comprising detecting, by computer, one or more pairs of non-adjacent DNA sequences to which are bound RNA molecule comprising two RNA sequences.

Claim 21 is directed to a method of identifying DNA sequences that control the expression of different collections of genes in a genome comprising detecting, by computer, changes in connectron behavior in the genome as a function of changes in the sequence of the genome.

Claim 25 recites the method of changing a given genome comprising modifying, by computer, the connectron organization therein.

Claim 26 is directed to:

a method of detecting, by computer, connectron control and target sequences in a given genome comprising:

determining the base composition of said

genome,

determining one or more sites of control

sequence organization and/or

determining one or more sites of target

application.

I agree that the skilled practitioner would turn to the description and drawings provided in the application for guidance in using the claimed invention. The specification provides a detailed roadmap for practicing the invention by one skilled in the art. Referring specifically to the specification and drawings, the introduction at pages 1 - 3 provides a basic description of connectron structure. Figures 1 - 3 are taken from the text by Alberts et al. entitled "The Molecular Biology of the Cell." Pages 3 - 25. Pages 26 - 36 provides a detailed description of a connectron structure. Page 31, the detailed description of the invention, provides a descriptive analysis of the flow diagrams utilized in the computer analysis of connectrons in any given genome.

Ten samples of connectrons found by computer mediation are set out in the specification. Pages 39 - 56 give an example of a prokaryote connectron - E. coli. I have considered this example as

well as all examples given against the backdrop of the Examiner's contention that the description lacks clear evidence of the connectron symmetries as related to gene expression and in my opinion that the skilled practitioner would not have any difficulty in practicing the invention from these descriptions for the following reasons:

(1) The flowcharting conventions used are typical of those used to present computer algorithms. Together, they provide all the detail required for a complete implementation.

(2) A wide variety of computer languages could be used to implement the algorithm. Any procedural third generation language could be used.

(3) These skills are well within the competence of even journeyman programmers using languages such as Fortran, Cobol, PL-I, ALGOL< Pascal, etc., as well as more modern languages such as C, C++, etc.

(4) Computers with the necessary performance and capacity are readily available for an amount that is well within the reach of

many home budgets, let alone the resources available to corporations and research institutions.

Finally, in conclusion, I disagree with the Examiner's contention that the trial and error experimentation required to practice the invention amounts to undue experimentation for the following reasons:

(1) As stated earlier, the algorithms presented are straightforward and complete.

(2) No experimentation whatsoever is required. Implementing the algorithms is a routine exercise in program design, coding and debugging. Running them is simply a matter of obtaining the organism-specific genomes and allowing the computer programs to go to work on them.

(3) The only part of the activity that could conceivably be referred to as "experimenting" is the investigation into available bioinformatics resources, such as the syntax and semantics of the resources provided by, for example, that National Library of Medicine's National Center for Biotechnology Information (NCBI). It is clear that in this context, having a ready understanding of this

Serial No. 09/866,925.....Page 14

information is a reasonable characteristic of one who could be called "skilled in the art."

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Date: 6/5/2003

James V. Oberthaler
(Signature)
Name: JAMES V. OBERTHALER



JAMES V. OBERTHALER
Principal Research Specialist

RECEIVED

JUN 11 2003

TECH CENTER 1600/2900

SUMMARY OF EXPERIENCE

- Hands-on manager, skilled in oral and written communications, analysis, problem solving, and strategic planning
- Member of Federal Senior Executive Service with significant experience in managing a wide range of development and operational/support activities, developing and integrating systems in a variety of hardware and software platforms
- Information technology policy making at senior levels in the Federal Government, responsibility for oversight and review of major information systems.
- Led major efforts in systems quality assurance, configuration management, information systems security, strategic planning, and systems operation.
- Strong background in systems design and development in a broad variety of architectures and programming languages.

PROFESSIONAL EXPERIENCE

Independent Consultant

2000 – present

Provides systems engineering and technical assistance to the National Cancer Institute's Center for Bioinformatics in developing enterprise-wide scientific information systems. The systems include

- The *Enterprise Vocabulary System (NCI-EVS)*, an on-line thesaurus that provides a common vocabulary for scientific and administrative personnel throughout the Institute; and,
- The *Intramural Research Directory*, a comprehensive system containing information describing research investigations conducted by NCI employees.

Reviews work done by other contractors; conducts analyses leading to improved business processes, coordinates implementation, ongoing operations, EVS project governance activities, etc.

Project Leader, Synectics for Management Decisions, Inc.

1994 – 2000

Managed projects at the Bureaus of the Census, Department of Commerce and the Substance Abuse and Mental Health Services Administration, DHHS

U.S. Patent and Trademark Office (PTO)

1988 – 1994

Director for Systems Engineering and Evaluation (1988–1991)

Monitored the work of the Automated Patent System (APS) Integration Contractor in designing and building APS components; improved the contractor's systems for quality assurance, configuration management, component error-handling, and reporting. Conceptualized, developed a comprehensive requirements document and detailed implementation plans for PTOnet, PTO's campus area network..

Director of Central Computer Operations (1991–1994)

Led an office of 130 government employees and numerous contract personnel in operating, maintaining, enhancing, and supporting the production information systems and basic computer systems infrastructure of the PTO.

**Deputy Assistant Secretary for Information Resources Management,
and Deputy Assistant Secretary for Management Analysis and Systems, DHHS** 1987 – 1988

Advised the DHHS Secretary and the Assistant Secretary for Management and Budget (ASMB) on issues and policies pertaining to the use of information resources.

Initiated projects of strategic importance: Department-wide electronic linkage for exchanging messages and files; a backbone campus-area network for DHHS headquarters and adjoining buildings; a standard Local Area Network for elements of the Office of the Secretary; an infrastructure of intelligent workstations and support for ASMB, including DHHS regional operations.

**Division of Computer Research and Technology (DCRT)
National Institutes of Health, DHHS** 1969 – 1987

Assistant Director, DCRT (1983—1987).

Assistant to Chief, NIH Computer Center (1977—1983).

Head, Systems Team, NIH Computer Center (1972—1977)

Directed systems development programming for the computer center.

Systems Programmer (1969—1972).

Participated in designing and writing the NIH Shared Spool System, a major modification to the IBM Operating System. This system was used by more than 100 installations worldwide and served as the model for IBM's Multi-Access Spool, a standard facility of IBM Operating Systems. Designed and wrote many enhancements to the WYLBUR text editing system. Developed DATASTOR, supporting peer-to-peer communications among computers.

Staff Programmer, IBM Federal Systems Division 1968 – 1969

Led the programmers who developed display software for the FAA's Air Traffic Control System.

Health Services Officer, USPHS, CCB, DCRT, NIH 1966 – 1968

Diagnosed system problems and effected repair or circumvention; developed modifications to the IBM operating system and did utility programming.

Senior Associate Programmer, IBM Federal Systems Division 1962 – 1966

Performed system and utility programming for Air Force Project 473L; designed and wrote the part of the system control program that allowed application programs to allocate and access disk storage.

TECHNICAL SUMMARY

IBM computers including 3090/200, large IBM computer centers; Unisys A-15, A-16; Amdahl 5990-1100; IBM and Macintosh Personal Computers. OS/MVS, TSO, WYLBUR, NIH Shared Spool.

Local Area Networks (PTOnet and the NIH Campus Network), PL-1, COBOL, IBM S/360 Assembler Language, variety of software on Apple Macintosh and IBM PC: Excel, Word, WordPerfect, PowerPoint, Visio, Mac System 9 and X, MacProject, Microsoft Project, Adobe Photoshop.

EDUCATION

BS Mathematics, Cum Laude, Holy Cross College, 1962. Graduate work in Mathematics, The American University, 1969-1970. MBA, Southern Illinois University at Edwardsville, 1978.

**ALGORITHMIC DETERMINATION OF FLANKING DNA
SEQUENCES THAT CONTROL THE EXPRESSION OF SETS
OF GENES IN PROKARYOTIC, ARCHEA AND EUKARYOTIC
GENOMES**

5

Reference to Related Application

The present application is the subject of Provisional
Application Serial No. 60/208,650 filed June 2, 2000
entitled ALGORITHMIC DETERMINATION OF CONNECTRONS FOR THE
10 HIGH LEVEL REGULATION OF GENE EXPRESSION.

Introduction

RNA introduced into a cell by a virus is now known to
trigger a cellular defense mechanism known as post-
transcriptional gene silencing (PTGS). If the viral RNA
15 sequence matches a sequence within the cell's genome the
associated genes are turned off or silenced. This
phenomenon is also called 'RNA interference' or RNAi. A
single-stranded RNA can interact with another single-
stranded RNA (known as antisense RNA). The single-
20 stranded RNA can also form a triple-stranded complex with
double-stranded DNA. This triple-stranded complex is
known as a Hoogsteen helix. This patent application
shows how two specific adjacent RNA single-stranded
sequences (called C1 and C2 - for Control Sequence 1 and
25 Control Sequence 2) interact with two distant double-
stranded DNA sequences (called T1 and T2 - for Target
Sequence 1 and Target Sequence 2) to form a tetradic
relationship which is called a "connectron". The two
distant DNA double-stranded sequences (T1 and T2) must be
30 on the same chromosome in a genome and they must be

between about 1kb and 105kb of each other. The adjacent single-stranded RNA sequences (C1/C2) can be on the same or different chromosome as the T1 and T2 sequences. The C1 sequence is identical to the T1 sequence and the C2 sequence is identical to the T2 sequence. The connectron acts to stabilize the double-stranded DNA by allowing 30nm chromatin particles to form. Genes that lie between the T1 and T2 sequences when wrapped up in 30nm chromatin particles are not open to promotion and expression. The connectron (i.e. the tetradic relationship between the T1-T2 sequences and C1/C2 sequences) provides a general explanation for PTGS. A connectron can be implemented by RNA sequences, PNA (Peptide Nucleic Acid) sequences or by a zinc-finger DNA Binding Protein (DBP) specific to the T1 and T2 sequences.

Characteristically the adjacent C1/C2 sequences lie in the 3'UTR of a gene. The T1 and T2 sequences do not lie within the translated region of any gene. These sequences "surround" one or more genes. There are, however, T1 and T2 sequence pairs that surround one or more C1/C2 sequences that are not 3'UTR to any gene. These are called "geneless connectrons". There may be promoter sequences that cause the transcription of these 3'UTR sequences.

A computer-based algorithm that is similar to the algorithm used in the US Patent 6,205,404 has been developed to determine the connectron structure of any genome. This algorithm determines the existence of all the connectrons in the genomic DNA. Connectrons exist in prokaryotes, archaea, single-celled eukaryotes, multi-celled eukaryotes, plants and higher animals. Connectron relationships exist between prokaryotes and their

plasmids. The geneless connectrons provide a possible mechanism for forming a hierarchy of gene expression control that will produce an understanding of cell differentiation and tissue development.

- 5 Each connectron is a unique tetrad of sequences. Each connectron changes the expression of the genes between the T1 and T2 sequences. The C1 sequence (which is equivalent to the T1 sequence) and the C2 sequence (which is equivalent to the T2 sequence) are determined by the
10 invention described in this patent application. In general, the tetrad of connectron sequences can be patented because the structure of matter is known and the function of specific gene expression modulation is also known. Gene expression modification can be produced by
15 introducing antisense RNA or PNA to interact C1/C2 RNA sequences or zinc-finger DBPs to interact with the T1 and T2 sequences. Using connectrons it will be possible to modify cellular and tissue behavior in a very general manner.
- 20 Examples will be given from different genomes to illustrate that the connectron is a perfectly general and universal concept.

Definitions

25

Double stranded DNA - Watson and Crick showed in 1953 that DNA naturally forms a double-stranded helix. A typical double stranded sequence is

- 30 5' -TAGAGGAGTACCAC-3'
3' -ATCTCCTCATGGTG-5'

Hydrogen Bond - The force between a hydrogen atom and another heavier atom such as Oxygen (O), Nitrogen (N), Phosphorus (P), or Sulfur (S).

5

Positive strand - The positive strand is normally represented 5' to 3' running left to right as in

5'-TAGAGGAGTACCAC-3'

10

Negative strand - The negative strand is normally represented 5' to 3' running right to left as in

3'-ATCTCCTCATGGTG-5'

15

Single stranded RNA - Either the positive or the negative strand of the double-stranded DNA can be transcribed by the polymerase. In RNA U replaces T.

20 RNA of positive strand sequence 5'-UAGAGGAGUACCAC-3'
RNA of negative strand sequence 5'-GUGGUACUCCUCUA-3'

Antisense RNA - The antisense strand of any RNA sequence is the compliment sequence

25

RNA sequence 5'-UAGAGGAGUACCAC-3'
Antisense RNA sequence 3'-AUCUCCUCAUGGUG-5'

30 Triple Strand Helix - The RNA sequence of a RNA/DNA triple-strand complex is the same as the positive strand of the DNA

DNA positive strand 5'-TAGAGGAGTACCAC-3'

DNA negative strand 3'-ATCTCCTCATGGTG-5'
RNA strand 5'-UAGAGGAGUACCAC-3'

Promoter - Any region of DNA, that binds proteins which
5 engage the polymerase transcription mechanism.

TATA Box - A region near the 3' end of a promoter with
the sequence TATA.

10 mRNA - The RNA produced from the DNA by the polymerase as
a result of transcription

Start of transcription - The 3' end of a promoter where
the polymerase mechanism begins to transcribe DNA into
15 mRNA.

Exon - Any region of mRNA which is used to code for
proteins

20 Intron - Any region of mRNA lying between two exons which
is not used to code for proteins. The introns are edited
out of the initial RNA transcript to form the mature
mRNA.

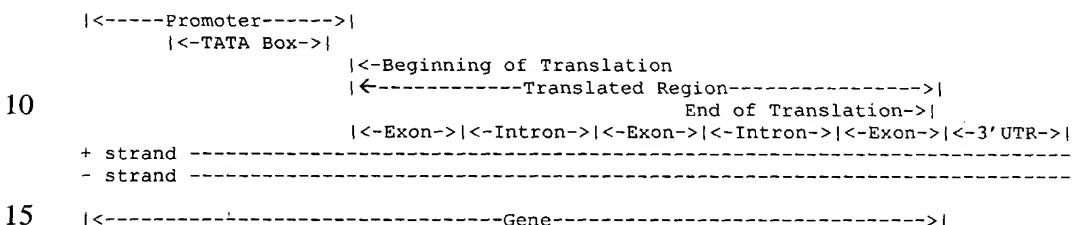
25 3' UTR - The untranslated 3' end of an mRNA is beyond the
end of the last exon. A stop codon in the mRNA causes
the ribosome to stop the translation of mRNA into
protein.

30 End of translation - The 3' end of the 3'-most exon.

Translated region - Any collection of exons and introns.

Gene - Any DNA region that codes for a protein. Introns do not occur in prokaryotic genes and they sometime fail to occur in eukaryotic genes. A typical model of a gene is

5



Positive strand gene - Any gene in which the features run 5' to 3' on the positive strand

20 Negative strand gene - Any gene in which the features run 5' to 3' on the negative strand

C1 sequence - Any positive or negative strand DNA sequence of 20 bases or more.

25 The C2 sequence must occur in the same chromosome as the C1 sequence.

C2 sequence - Any positive or negative strand DNA sequence of 20 bases or more.

30 The C1 sequence must occur in the same chromosome as the C2 sequence.

C1/C2 - Any positive or negative strand DNA sequence of 40 or more bases such that the C1 sequence is adjacent to
35 the C2 sequence

T1 sequence - Any positive or negative strand DNA sequence of 20 bases or more that is on the same

chromosome as the T2 sequence. The T1 and T2 sequences must be between about 1kb and 105kb apart.

5 T2 sequence - Any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence. The T2 and T1 sequences must be between about 1kb and 105kb apart.

10 Last exon gap or Gap-Distance - The number of bases between the end of transcription and the beginning of the C1/C2 sequence. In prokaryotes and single-celled eukaryotes this gap can range from no bases to 500 bases. In multi-celled eukaryotes the gap can be as large as 10,000 bases.

15 Poly-adenylation signal - A number of Adenosine (A) bases are added to the mRNA at the end of the 3'UTR.

20 Possible Connectron - Any set of T1, T2 and C1/C2 sequences such that the C1 sequence is identical to the T1 sequence and the C2 sequence is identical to the T2 sequence. The promoter of some gene causes the mRNA of the gene to be expressed. The mRNA is edited to eliminate the introns. The whole mRNA including the 3'UTR can move about in the cell or the nucleus of the cell. The C1/C2 RNA that is part of the 3'UTR moves to the T1 and T2 DNA sequences. A triple-stranded complex of the DNA and the RNA forms such that the C1 sequence forms hydrogen bonds with the T1 sequence and the C2 sequence forms hydrogen bonds with the T2 sequence. Because the C1 sequence is adjacent to the C2 sequence, the T1 sequence is brought physically close to the T2 sequence. This produces a loop of between about 1kb and

105kb in the DNA. Histone proteins reduce the length of the DNA by binding 200 bases. Histone/DNA complexes form six-fold symmetry chromatin assemblies. The diameter of the chromatin assemblies is approximately 30nm.

5

Real Connectron - Any Possible Connectron which is within the Gap-Distance of some gene

10 Homologous connectron - The T1 sequence and the T2 sequence are on the same chromosome as the C1/C2 sequence

Heterologous connectron - The T1 sequence and the T2 sequence are on a chromosome different from chromosome of the C1/C2 sequence

15

Permanent connectron - Any C1/C2 sequence, which is 3' UTR to some gene that is not surrounded by any T1 and T2 sequence pairs

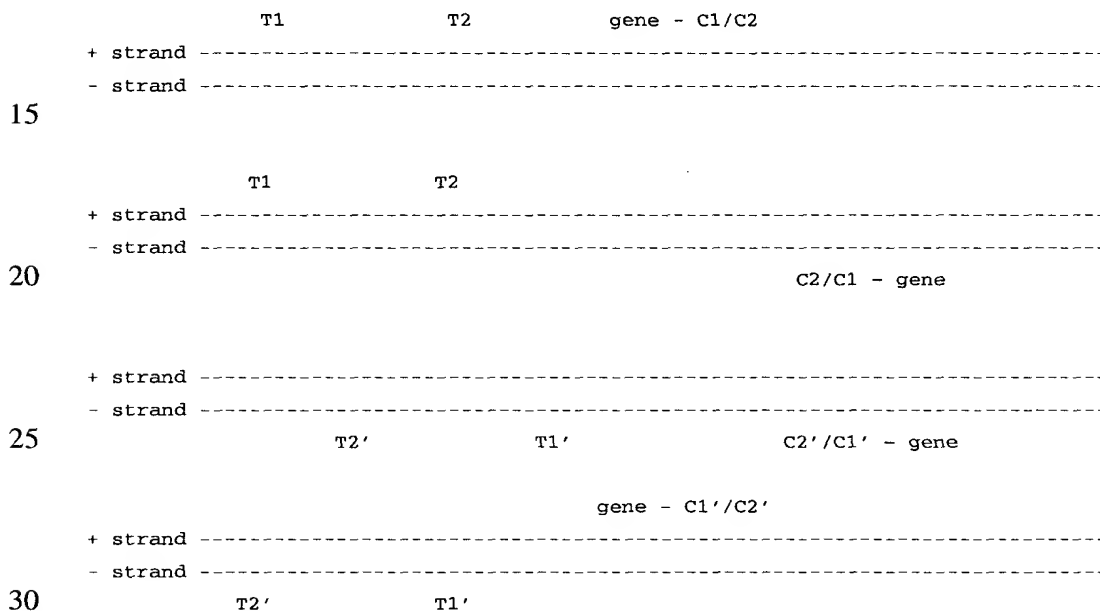
20 Transient connectron - Any C1/C2 sequence, which is 3' UTR to some gene that is surrounded by one or more T1 and T2 sequence pairs

25 Self-limiting connectron - Any C1/C2 sequence which is 3'UTR to some gene that is surrounded by the T1 and T2 sequences such that C1=T1 and C2=T2

30 Geneless connectron - Any C1/C2 sequence which is not 3'UTR to some gene but is surrounded by some T1 and T2. A promoter may lie 5' to the C1/C2 sequence.

Bidirectionality of Connectron Excitation - A C1/C2 short loop on one strand selects a T1-T2 long loop pair on the

same or the opposite strand. The C1/C2 short loop has a complementary C1'/C2' sequence on the opposite strand. Similarly the T1-T2 long loop pair has a complementary long loop pair T1'-T2'. Wherever a C1/C2, T1-T2 tetrad
 5 exists there is a complementary C1'/C2', T1'-T2' tetrad. The C1/C2 short loop can be transcribed as a 3'UTR to a gene on the same strand. The C1'/C2' short loop which is on the strand opposite to the C1/C2 short loop can also can be transcribed as a 3'UTR to a gene on the same
 10 strand. There are four possible models of action



Of course, the short loops and the long loops do not have to be on the same chromosome.

35 Hierarchy of connectron action - When a C1/C2 is expressed it forms a T1-T2 loop by forming a connectron. The C1/C2 sequence does not have to be on the same chromosome as the T1 and T2 sequences. This provides a way of causing interaction between chromosomes. When the
 40 T1-T2 loop forms, any genes in that loop region which had

been expressing C1/C2 sequences in their 3'UTRs, now
cease expressing the C1/C2 sequences. The connectrons
formed by these C1/C2 sequences will cease to exist after
some time thus opening up the genes inside the respective
5 T1-T2 loops to expression. The hierarchy of connectron
action is alternates between repression and expression.
The connectron hierarchies can be of any depth.

One-to-Many connectron action - One C1/C2 sequence can
10 form connectrons in many different places on many
different chromosomes. The only requirement is that
C1=T1 and C2=T2. This makes it possible for one
expression event to control the expression of many genes
on different chromosomes.

15 Many-to-One connectron action - C1/C2s that come from
many different places on many different chromosomes can
form a connectron for a specific T1-T2 sequence pair.
The only requirement is that C1=T1 and C2=T2. This makes
20 it possible for many different expression events to
control the expression of one set of genes on a
particular chromosome.

Many-to-Many connectron action - The arrangement of
25 C1/C2s and T1-T2s across chromosomes can form a complex
web of gene expression control relationships.

Percentage of the Genome Regulated by Connectrons - Since
the connectrons for a sequenced genome can be calculated,
30 the percentage of the genome that is open to connectron
regulation can be known.

Emergent Property - The network of connectrons in any genome emerges from a knowledge of the complete DNA sequence of the genome. Because both the C1/C2 sequences and the T1-T2 sequences can be any place in the genome,
5 the whole genomic sequence must be known before all the connectrons can be determined.

Paradigm Shift - For the past fifty years since the discovery by Watson and Crick of the double-helical
10 nature of DNA, the reigning paradigm for scientific discovery has been the study of one gene and its effects on the behavior of a cell. The advent of genomic sequencing and this invention of connectrons that emerge from the whole genome will produce a shift in the way
15 scientists view biological systems and the way they formulate and execute experiments. The many-to-many relationships between the connectrons means that there are many ways in which the expression of a set of genes can be modulated. The multiplicity of control pathways
20 means produces a system stability that makes it possible for biological systems to be stable for long periods of evolutionary time. The thinking that goes into formulating scientific experiments will have to change to accommodate the changes in understanding that will be
25 induced by the application and extension of this patent application.

Hierarchy of DNA Structuring - The DNA of a cell's genome is structured in a hierarchy of six levels. Figures 1, 2
30 and 3 have been adapted from The Molecular Biology of the Cell by Alberts, Bray, Lewis, Raff, Roberts and Watson [third edition pages 354, 345 and 348]. As shown in figure 1, the double stranded DNA is level 1. The

double-stranded DNA is wrapped around histone proteins to form a chromatin particle that is level 2 of the hierarchy. Level 2 is described as "beads-on-a-string" in figure 1. The chromatin particles are packed in a six-fold symmetry as shown in figure 2a and figure 2b. These six-fold assemblies have a diameter of 30 nm. Each 30 nm assembly contains from 18 (i.e. $6 * 3$) to 30 (i.e. $6 * 5$) chromatin particles. The 30 nm assemblies aggregate into large loops which range in length from 5,000 bases to 100,000 bases of DNA. The size of these large loops as shown in figure 1 is approximately 300 nm. These large loops constitute level 4 of the structuring hierarchy. As shown in figure 1, level 5 of the DNA structuring hierarchy many large loops are condensed to form a structure which is approximately 700 nm in diameter. The complete chromosome that constitutes level 6 of the hierarchy is composed of two very long sections of level 5 DNA.

Model of Chromatin Structure - The level 4 structure of DNA as shown in figure 1 ranges in length from 5,000 to 105,000 bases of DNA. Figure 3 shows that proteins are thought to connect portions of the long loops formed by the 30 nm particles to form a chromosome axis. These condensed long loops are described as chromomeres in The Molecular Biology of the Cell.

Prior Art

The chromomere model of DNA structuring was presented by N. A Resnik, et al.[1] and is based on electron microscopic data. There are more recent papers studying

a variety of genomes with electron microscopy but no equivalent study of chromomeres has been done on a fully sequenced genome.

5 A recent News Feature in Nature by T. Gura [2] described the discovery of post-transcriptional gene silencing in which viral RNA interacts with the transcribed RNA of the cell to silence the expression of genes. This article describes experiments in *C. elegans* and *D. megalomaster* in which RNA that is complementary to mRNA introduced
10 into a cell. This "antisense" RNA has the effect of turning off the expression of one or more genes. The introduced complementary RNA produces an "RNA interference" called RNAi.

15 Thomas Werner and his colleagues at Genomatix in Munich, Germany have developed an approach to understanding what they call "Matrix Attachment Region" (MAR). Figure 5 shows their interpretation of the structure of DNA surrounding a gene. The following description of the MAR is copied from the Genomatix web site

20 "Matrix Attachment Regions (MARs) MARs are sequence regions that are responsible for the attachment of genomic DNA to the nuclear matrix or scaffold. Transcription absolutely requires anchorage of genomic
25 DNA to the nuclear matrix.

Functional features of MARs:

30 Anchoring of regulatory elements like promoters and enhancers to the nuclear matrix.

Ensuring long term activity of promoters and enhancers in chromatin.

5 Insulation, rendering a functional domain insensitive to position effects.

Genomatix is conducting a research project to define and detect MARS by computer-analysis."

Brief Description of the Objects of the Invention

5 An object of the invention is to provide a method of identifying DNA sequences that control the expression of different collections of genes in a genome comprising, detecting selected DNA sequences adjacent to some genes excluding exons and introns.

10 An object of the invention is to provide a method of identifying DNA sequences that control the expression of different collections of genes comprising, detecting, by computer, one or more pairs of non-adjacent DNA sequences to which are bound to two RNA sequences.

15 An object of the invention is to provide a method of identifying DNA sequences that control the expression of different collections of genes in a genome comprising detecting changes in connectron behavior in the genome.

20 An object of the invention is to provide a method of modifying the expression of different gene collections in a genome, comprising detecting changes in connectron behavior as a result of an exogenous stimulus.

25 An object of the invention is to provide a method of detecting where and when new genes are being integrated into a host genome comprising detecting the connectrons in said host genome.

30

An object of the invention is to provide a method of detecting the expression effect of different gene collections in a given body comprising detecting the back and forth flow of connectrons between the chromosomes thereof.

An object of the invention is to provide a method of modifying a given body comprising modifying the connectron organization therein.

10

An object of the invention is to provide a method of detecting connectron control and target sequences in a given genome comprising:

15 determining the base composition of said genome,
 determining one or more sites of control sequence organization, and/or
 determining one or more sites of target application.

20 An object of the invention is to provide a method of determining the response of a cell in any tissue to changes in the cell's environment and/or genetic composition comprising providing a complete genomic DNA sequence for the organism and determining the effect of
25 changes in connectrons due to application of a given exogenous stimulus to the genome.

An object of the invention is to provide a method of determining in prokaryotes, archea, single-celled
30 eukaryotes and multi-celled eukaryotes, the tetradic relationship $T1=C1$ and $T2=C2$ where $T1$ and $T2$ are DNA sequences 20 or more bases in length, where the $C1$ sequence is adjacent to the $C2$ sequence, where the $T1$ and

T2 sequences are on the same chromosome, and where the C1/C2 sequences are on the same chromosome as T1 and T2 or where the C1/C2 sequences are on a chromosome different from T1 and T2, wherein:

5

C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

10

C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

15

C1/C2 - any positive or negative strand DNA sequence of 40 or more bases such that the C1 sequence is adjacent to the C2 sequence,

20

T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart, and

25

T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart.

30

An object of the invention is to provide a method of determining in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes, the connectron relationship that permits many different C1/C2 short loops to control the existence of a T1-T2 long loop and

wherein said C1/C2 short loops can be on the same chromosome or on different chromosomes from the T1-T2 long loop, wherein:

5 C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

10 C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

15 C1/C2 - any positive or negative strand DNA sequence of 40 or more bases such that the C1 sequence is adjacent to the C2 sequence,

20 T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart, and

25 T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart.

30 An object of the invention is to provide a method of determining in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes, the connectron relationship that permits one C1/C2 short loop to control the existence of many T1-T2 long loops, the C1/C2 short

loop can be on the same chromosome or on different chromosomes from the T1-T2 long loops, wherein:

5 C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

10 C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

15 C1/C2 - any positive or negative strand DNA sequence of 40 or more bases such that the C1 sequence is adjacent to the C2 sequence,

20 T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart, and

25 T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart.

30 An object of the invention is to provide a method of determining in the connectron relationships between prokaryotes and their plasmids wherein said connectrons implement a control mechanism between the two genomes that makes it possible from them to form a symbiotic relationship, and in the case of *D. radiodurans* the relationship is not symmetric, and the *D. radiodurans*

genome sends C1/C2 short loops to the MP1 plasmid,
wherein:

5 C1 sequence - any positive or negative strand DNA
sequence of 20 bases or more, the C2 sequence must
occur in the same chromosome as the C1 sequence,

10 C2 sequence - any positive or negative strand DNA
sequence of 20 bases or more, the C1 sequence must
occur in the same chromosome as the C2 sequence,

15 C1/C2 - any positive or negative strand DNA sequence
of 40 or more bases such that the C1 sequence is
adjacent to the C2 sequence,

20 T1 sequence - any positive or negative strand DNA
sequence of 20 bases or more that is on the same
chromosome as the T2 sequence, the T1 and T2
sequences must be between about 1kb and 105kb apart,
and

25 T2 sequence - any positive or negative strand DNA
sequence of 20 bases or more that is on the same
chromosome as the T1 sequence, the T2 or T1
sequences must be between about 1kb and 105kb apart.

An object of the invention is to provide a method of
determining that connectron relationships that exist in
plant and higher animals.

30

An object of the invention is to provide a method of
determining in prokaryotes, archea, single-celled
eukaryotes and multi-celled eukaryotes, the connectron

relationship that permits one C1/C2 short loop to control the existence of one or more T1-T2 long loops without being subject to any expression controls other than those of the gene to which the C1/C2 is 3'UTR, wherein:

5

C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

10

C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

15

C1/C2 - any positive or negative strand DNA sequence of 540 or more bases such that the C1 sequence is adjacent to the C2 sequence,

20

T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart,

25

T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart, and

30

3'UTR - untranslated 3' end of an mRNA is beyond the end of the last exon, a stop codon in the mRNA causes the ribosome to stop the translation of mRNA into protein.

An object of the invention is to provide a method of determining in prokaryotes, archea, single-celled eukaryotes and multi-celled eukaryotes, the connectron relationship that permits one C1/C2 short loop to control
5 the existence of one or more T1-T2 long loops such that this C1/C2 short loop is itself subject to expression control by another T1-T2 long loop which surrounds it, wherein:

10 C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

C2 sequence - any positive or negative strand DNA
15 sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

C1/C2 - any positive or negative strand DNA sequence of 540 or more bases such that the C1 sequence is
20 adjacent to the C2 sequence,

T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2
25 sequences must be between about 1kb and 105kb apart, and

T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1
30 sequences must be between about 1kb and 105kb apart.

An object of the invention is to provide a method of determining in prokaryotes, archea, single-celled eukaryotes and multi-celled eukaryotes, the connectron relationship that permits one C1/C2 short loop to control
5 the existence of the T1-T2 long loop that surrounds it, wherein:

C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must
10 occur in the same chromosome as the C1 sequence,

C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,
15

C1/C2 - any positive or negative strand DNA sequence of 40 or more bases such that the C1 sequence is adjacent to the C2 sequence,

20 T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart, and

25 T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart.

30

An object of the invention is to provide a method of determining the connectron relationships that do not have any genes within the T1-T2 long loop, wherein:

T1 sequence is any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, and

5

T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, and the T2 or T1 sequences must be between about 1kb and 105kb apart.

10

An object of the invention is to provide a method of determining the geneless connectron relationship where one C1/C2 short loop controls the existence of many geneless T1-T2 long loops, wherein:

15

C1 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C2 sequence must occur in the same chromosome as the C1 sequence,

20

C2 sequence - any positive or negative strand DNA sequence of 20 bases or more, the C1 sequence must occur in the same chromosome as the C2 sequence,

25

C1/C2 - any positive or negative strand DNA sequence of 40 or more bases such that the C1 sequence is adjacent to the C2 sequence,

30

T1 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T2 sequence, the T1 and T2 sequences must be between about 1kb and 105kb apart, and

T2 sequence - any positive or negative strand DNA sequence of 20 bases or more that is on the same chromosome as the T1 sequence, the T2 or T1 sequences must be between about 1kb and 105kb apart.

Description of the Drawings and Tables

The above and other objects, advantages and features of the invention will become more apparent when considered with the following specification and accompanying drawings and tables wherein:

Figure 1 DNA is structured in six levels of increasing condensation. Double stranded DNA is level 1. Two turns of DNA are wrapped about each chromatin particle at level 2. The chromatin particles which each containing 200 base pairs form into 30 nm particles at level 3. The 30 nm particles form into large loops with an approximate dimension of 300 nm at level 4. Metaphase chromosomes form a condensed structure with an approximate dimension of 700 nm at level 5. An entire metaphase chromosome has a width of approximately 1400 nm at level 6. The large loops at level 4 of the DNA structuring are thought to have between 20,000 (20 kb) and 100,000 (100 kb) base pairs.

The Molecular Biology of the Cell by Alberts, Bray, Lewis, Raff, Roberts and Watson, 3rd. ed., Garland Publishing, Inc., New York, 1994, p. 354

Figure 2 (a) Chromatin DNA forms into a six-fold symmetry 30nm particles.

5 (b) The six-fold symmetry 30nm particles form a linear chain with a varying number of repeat units.

The Molecular Biology of the Cell by Alberts, Bray, Lewis, Raff, Roberts and Watson , 3rd. ed. , Garland Publishing, Inc., New York, 1994, p. 345

Figure 3 Long loops of 30nm particles are thought to be closed at the bottom of the loop by proteins.

15 The Molecular Biology of the Cell by Alberts, Bray, Lewis, Raff, Roberts and Watson, 3rd. ed. , Garland Publishing, Inc., New York, 1994, p. 348

Figure 4 (a) Transcription and Editing. (b) Movement of the RNA through the Nucleus. (c) Connectron Formation

Figure 5 Overview of schematic organization of a typical transcriptionally active chromosomal loop.

Table 1 Connectron Properties for Prokaryotic, Archea and Eukaryotic Genomes

Table 2 Yeast Inter-Chromosomal Connectron Distribution

Figure 6 Genome size plotted as a log-log function of
the Number of Connectrons

Figure 7 Number of Sequence Instances plotted as a
function of the Number of Fragments

5 Figure 8 Level 0 - The overall view of the algorithm

Figure 9 Level 1 - Process Flow of the Algorithm

Figure 10 Level 2a - two pages - Process Genome into
Blocking Fragment File

10 Figure 11 Level 2b - three pages - Compute the
Connectrons for a Genome

Figure 12 Level 2c - two pages - Analyze Possible
Connectrons

Figure 13 Level 3a - Setup Genome Usage Memory

15 Figure 14 Level 3b - Find DBP-Size Blocking File for T1-
Window

Figure 15 Level 1 - Find DBP-Size Blocking File for T2-
Window

Figure 16 Level 2a - two pages - Find C1/C2 Entries

20 Figure 17 Level 2b - two pages - Scan Genome Usage Memory
for Potential Connectrons

Description of the Invention

A connectron is a relationship among four DNA sequences. Each sequence must be at least 20 bases long. There is a
5 report by Sharp and Zamore [3] that RNA sequences of "about length 25" are important as sources of RNAi. 27 bases were actually used as the minimum length of each of the sequences. The T1 sequence is on one strand of some chromosome in a genome. The T2 sequence is on the same
10 strand of the same chromosome as the T1 sequence. The T1 and T2 sequences (which are each at least 20 bases in length) must be at least 5,000 bases distant from each other but they can not be more than 105,000 bases distant from each other. The C1 sequence and the C2 sequence
15 (which are each at least 20 bases in length) are adjacent to each other on some strand of some chromosome in the genome. The C1/C2 sequences - called the "short loop" - can be on the same strand as the T1 and T2 sequences or they can be on the opposite strand. The C1/C2 sequences
20 of the short loop can be on the same chromosome as the T1 and T2 sequences but they can also be on a different chromosome in the genome. When a genome has only one chromosome, then the point is moot. Many genomes, of course, have several chromosomes. The C1 sequence is
25 identical to the T1 sequence and the C2 sequence is identical to the T2 sequence.

The C1/C2 sequence must be on the same strand as a gene, either be directly adjacent to the gene (i.e. a gap of 0
30 bases) for prokaryotic genomes or at this time be within 10,000 bases for eukaryotic genomes. The size of the gap between the end of the gene and the beginning of the

C1/C2 sequence is a variable. The C1/C2 short loop is expressed as the 3'UTR (Un-Translated Region) of the gene. In the case of prokaryotic genes that do not normally have introns, the whole mRNA becomes the active species for connectron formation. In the case of eukaryotic genes, the whole transcript is the active species for connectron formation upon editing of the transcript to eliminate the introns. The whole transcript then can move about in the cytoplasm of prokaryotic cells or the nucleus of eukaryotic cells. Since the C1 sequence is equivalent to the T1 sequence and the C2 sequence is equivalent to the T2 sequence, the C1 RNA can form a Hoogsteen triple-stranded RNA/DNA/DNA helix with the double-stranded T1 sequence. Similarly the C2 RNA can form a Hoogsteen triple-stranded RNA/DNA/DNA helix with the double-stranded T2 sequence. Because the C1 sequence and the C2 sequence are adjacent to each other, the C1/T2 RNA/DNA/DNA Hoogsteen triple helix is brought into physical adjacency to the C2/T2 RNA/DNA/DNA Hoogsteen triple helix. RNA/DNA/DNA hybrid helices are the most stable form of triple helix. RNA double helices, DNA double helices, RNA triple helices and DNA triple helices are all significantly less stable than a RNA/double-stranded DNA triple helix. The stable physical adjacency of the two triple-stranded Hoogsteen helices ensures that the long loop of double-stranded DNA between the T1 sequence and the T2 sequence can then be structured into 30 nm chromatin particles as shown in level 4 of figure 1. The genes on either strand of the DNA between the T1 sequence and the T2 sequence when they are structured into the 30 nm chromatin particles are not open to promotion and expression.

The tetradic relationship between the T1 and T2 sequences that form the long loop and the C1/C2 sequences that form the short loop are called a connectron. The name "connectron" was suggested by J. David Rawn Ph.D. of
5 Towson University. A connectron is possible if the T1, T2, C1 and C2 sequences exist. A connectron is real if the C1/C2 short loop sequence is adjacent to an expressible gene. If the expression of the adjacent gene is inside one or more T1 - T2 long loops then this
10 connectron is said to be transient. If the adjacent gene is not inside any possible T1-T2 long loop then the connectron is said to be permanent. If a connectron is inside of a T1-T2 long loop that has the same sequences (i.e. T1 is really equal to C1 and T2 is really equal to
15 C2) then the connectron is said to be self-limiting. This is true because once the C1/C2 sequence is expressed it forms the T1-T2 long loop that then shuts off the expression of the gene adjacent to the C1/C2 sequence. Self-limiting connectrons can also be called "spike"
20 connectrons since they generate a short-duration spike of the C1/C2 short loop sequence. If a T1-T2 long loop does not contain any genes but it contains C1/C2 short loop sequences then this type of connectrons is said to be geneless. The C1/C2 short loops within a geneless T1-T2
25 long loop can, of course, control the expression of genes.

The physical existence and lifetimes of the connectrons must be proved by molecular biological experimentation.
30 This physical experimental process, however, is logically quite separate from the computational experimentation that have been conducted from June of 1999 to May of 2001. The computational search for the existence of

connectrons has been extremely positive. These computations have shown that connectrons exist in prokaryotes, in archea, between prokaryotes and their plasmids, in single-celled eukaryotes, in multi-celled eukaryotes, in plants, in higher animals and in humans. All of these features and properties are described in the claims section that follows.

The connectron invention is very powerful. It depends only on sequence equivalency. The minimum length of the four sequences seems to be about 20 bases. In the calculations shown in this patent application, 27 bases have been used as a minimum. The Nature News Feature [1] says that other scientists have found RNA sequences of length about 25 that have interesting gene silencing properties. The Nature article does not give any mechanism. Because of my algorithm and its use on a variety of genomes, this patent application provides the computational proof that a particular mechanism is highly probable. The connectron invention provides an explanation for how communication occurs with a chromosome as well as between chromosomes in genomes that have more than one chromosome. Since each T1-T2 long loop can contain one or more genes, the connectron invention provides a mechanism for turning on and turning off sets of genes simultaneously. In time, the connectron invention will provide an explanation for how differentiation of how one cell's behavior differs from the behavior of another adjacent cell. It is already clear from the computational experiments that have been made on *S. cerevisiae*, *C. elegans* and *D. megalomaster* that the number of geneless connectrons increases dramatically as evolution proceeds from single-celled eukaryotes (i.e.

S. cerevisiae) to 1,000 cell eukaryotes (i.e. C. elegans) to visible creatures (i.e. D. megalomaster). The extension of this evolutionary progress to plants (i.e. A. thaliana) for which only three chromosomes are
5 sequenced and humans (i.e. H. sapiens) for which only one chromosome is completely sequenced. Although the complete human genome was published in Nature and Science in February of 2001, the NIH-sponsored genomic sequencing results are available for about 1/3 of the bases in the
10 whole genome. The human genomic sequence determined by Celera Genomics, Inc. is available only by subscription. Table 1 shows how the genome size, the number of genes, the number of gene-containing and geneless connectrons and the percentage of genes controlled are related in
15 many different genomes.

The C1/C2 short loops originate on one chromosome. The T1-T2 long loops can be on the same or different chromosomes. Table 2 which is for yeast (S. cerevisiae)
20 is a square matrix of how many C1/C2 short loops on a given chromosome are sent to form T1-T2 long loops on other chromosomes. The diagonal of this matrix shows that many chromosomes send connectrons to themselves. The striking feature of this particular table is that
25 chromosome 6 only sends connectrons to chromosome 12 but that it receives connectrons from chromosomes 4,5,7,10,12,13,15 and 16.

Any tetrad of connectron sequences (i.e. the T1, T2, C1 and C2 sequences) as well as the fact of the adjacency of
30 the C1/C2 short loop sequence to the transcribing gene can be patented because the content of matter and the utility can be exactly described. The utility of a

connectron is that the T1-T2 long loop shuts off the expression of the genes that lie between the T1 sequence and the T2 sequence. In the case of geneless connectrons, the utility is of a higher level in that the
5 C1/C2 short loops contained in the higher-level geneless T1-T2 long loop, eventually form other lower-level T1-T2 long loops around a set of genes.

The invention of connectrons comes at a particularly
10 important time in biological discovery. The geneless connectrons make a many-to-many hierarchical control mechanism possible. It is already clear from the determination of the connectrons for *C. elegans* and *D. megalomaster* that there are as many or more geneless
15 connectrons than there are genes. It has been clear for some time that the number of genes in a genome is not particularly correlated with the size of the genome. Figure 6 shows that the size of a genome is roughly linearly correlated with the number of connectrons.

20 The connectron invention can be used to generate a model of behavior in any cell. The simulation of connectron behavior in different genomes will be the subject of another patent application.

25 The connectron invention provides for a rational exploitation of the information contained in the raw genomic DNA sequence by forming a hierarchy of relationships between geneless connectrons, transient
30 connectrons, permanent connectrons, self-limiting connectrons and the expression of genes.

THIS SHEET INTENTIONALLY LEFT BLANK.

Detailed Description of the Invention

The algorithm for the determination of connectrons in any genome or any genome fragment is represented in the following flow diagrams. The Level 0 diagram in figure 8 shows the general relationships in a digital computer. The central processor of the digital computer uses the computer program to take genome descriptors, the genomic DNA sequences and the tables of gene features to produce a file of blocking fragments and a file of the optimal connectrons for the genome. The printer serves to make hard copies of the files and this patent application. The level 1 diagram in figure 9 shows the three essential steps in the determination of connectrons. The genome is first processed into a blocking fragment file. Then the blocking fragments are used to compute the connectrons for the genome. Finally the potential connectrons are analyzed to determine if the C1/C2 sequences are in the 3'UTR of a gene. The level 2a diagram in figure 10 shows the steps required for the processing of the genome into a file of blocking fragments. The genomic DNA sequence is decomposed into 27-base frames for both the positive and negative strands. These fragments are written to the unsorted fragment file. The fragment file is then sorted is then read and formed into groups of equivalent sequences. The (.blk) file contains the sequence and a pointer to the (.gpstr) file which contains the pointers to the position of the fragments in the genomes. The position in the genome includes the chromosome number, the position in the chromosome and the strand (i.e. positive and negative). A sample of these files follows

Sample of the (.blk) file for *S. cerevisiae*

	27-base fragment (.gpstr) file	Number of instances	Pointer to
5	11111111111111111111111111111111	0	1
	1111111123244233313332443414	1	2
	111111141113443133314333341	2	4
10	111111232442333133324434141	1	5
	111111323311133323144423444	2	7
	111111332213331341414443413	2	9
	111111333444112343412323243	1	10
	111111333444113343412323243	9	19
15	111111411134431333143333414	2	21
	111111443223134142124434124	2	23
	111112223234344444443144442	2	25
	111112244123441122214421213	8	33
	111112311241114344334134431	2	35
20	111112324423331333244341414	1	36
	111112344232231344242234342	1	37
	111112433444244421144134211	1	38
	111112444311313442332142224	1	39
	111113131241131114424413231	1	40
25	111113143332344311113133411	1	41
	111113233111333231444234441	2	43

In fragments above 1=G, 2=C, 3=A, 4=T

30 Sample of the (.gpstr) file for *S. cerevisiae*

There are 16 chromosomes in *S. cerevisiae*

	Item	Chromosome	Position in Chromosome	Direction
35	1	0	0	0
	2	4	11137	1
	3	12	467619	1
40	4	12	458482	1
	5	4	11138	1
	6	12	465759	2
	7	12	456622	1
	8	1	219366	1
45	9	8	539978	1
	10	14	522451	1
	11	4	1099073	1
	12	4	1210003	1
	13	7	539068	1

	14	12	654136	1
	15	12	596455	1
	16	15	121016	1
	17	15	598127	2
5	18	16	847724	1
	19	16	59765	1
	20	12	467620	1
	21	12	458483	1
10	22	12	461657	1
	23	12	452520	1
	24	13	838006	1
	25	15	288270	1
	26	4	83593	1
15	27	4	992867	1
	28	6	162265	1
	29	7	845687	1
	30	10	531560	2
	31	15	282208	1
	32	16	860418	1
20	33	16	572308	1
	34	12	465992	1
	35	12	456855	1
	36	4	11139	1
	37	8	89343	1
25	38	4	10302	1
	39	1	19894	2
	40	16	9311	1
	41	10	735203	1
	42	12	465760	1
30	43	12	456623	1

In direction column above 1=positive strand,
2=negative strand

35

The level 2b diagram in figure 11 shows the computation of the connectrons. The genome descriptors consist of the number and length of the chromosomes. The algorithm uses an array that represents several facts about each base position in the genome. The level 3a diagram in figure 13 shows the setup of the Genome-Usage memory. The gene features are used to prevent the region of the genome that codes for proteins from being used for the connectron sequences (i.e. the T1s, the T2s, the C1s and the C2s). In the level 2a diagram of figure 10, the

45

algorithm steps through each chromosome and within each chromosome through each base position looking for acceptable T1-windows of 27 bases. A T1-window can be used to form a connectron relationship if there are two
5 or more instances of this fragment in the blocking fragment file. The computation in the level 3b diagram of figure 14 determines if the T1-window is acceptable or not. Once an acceptable T1-window is found, the algorithm (in the level 2a diagram of figure 10) looks
10 for acceptable T2-window positions that lie between 5,000 and 105,000 bases from the T1-window. The computation for determining acceptable T2-window positions is done in the level 3c diagram of figure 15. Once a pair of T1 and T2 window positions are found, the algorithm looks among
15 the instances of these T1 and T2 sequences for a pair of sequences C1 and C2 that lie within 200 bases of each other on the same chromosome. The computation for determining acceptable C1/C2 windows is shown in the level 3d diagram in figure 16. In the level 3e diagram
20 of figure 17 the Genome-Usage memory is scanned for the Possible-Connectrons. In the level 2c diagram of figure 12 the Possible-Connectrons are scanned to determine if the C1/C2 sequences are within the Gap-Distance of a gene on either the positive or the negative strands. The
25 Real-Connectrons are then written out in several different files including the descriptions in the claims section.

Examples

The algorithm for the determination of optimal connectrons has been applied to a number of different publicly available genomes. The connectron is a tetradic
5 relationship between four sequence elements - T1, T2, C1 and C2. The claims presented in this section are written by the program NearGene that implements the flow diagram Level 2c of figure 12. The examples are written a uniform type of English. Each example contains some or
10 all of the following elements

```
-----
Name of genome
Description of T1
15 Length of T1-T2 loop
The chromosome on which the T1-T2 loop exists
The identifier number within the genome of the T1
sequence
The T1 sequence
20 Description of T2
The identifier number within the genome of the T2
sequence
The T2 sequence
A list of genes whose expression is controlled by
25 the T1-T2 loop
The common names of the genes as obtained from the
NCBI gene feature file (.ptt)
A list of C1/C2 short loops whose expression if
controlled by the T1-T2 loop
30 The chromosome on which the C1/C2 short loop exists
The common name of the gene which expresses the
C1/C2 short loop as an RNA
```

- The sequence of the C1/C2 short loop
- A list of C1/C2 short loops that control the formation of the T1-T2 loop
- The chromosome on which the C1/C2 short loop exists
- 5 The common name of the gene which expresses the C1/C2 short loop as an RNA
- The sequence of the C1/C2 short loop
- The match between the C1/C2 sequence and the T1 sequence
- 10 The match between the C1/C2 sequence and the T2 sequence

The uniform descriptions make it possible to rapidly comprehend the specifics in each example.

- 15 When a sequence element is very long a series of four dots has been inserted between the beginning and ending sequence groups. A variable number of bases have been deleted.

20

Index for Connectron Samples

1. Connectrons occur in prokaryotes, archea, single-celled eukaryotes and multi-celled eukaryotes.
5
2. Many Connectrons control the expression of one set of genes in prokaryotes, archea, single-celled eukaryotes and multi-celled eukaryotes.
- 10 3. One connectron controls the expression of many sets of genes in prokaryotes, archea, single-celled eukaryotes and multi-celled eukaryotes.
- 15 4. Connectrons occur between prokaryotes and their plasmids.
5. Connectrons occur in plants and higher animals
- 20 6. Permanent connectrons exist in prokaryotes, archea, single-celled eukaryotes and multi-celled eukaryotes.
- 25 7. Transient connectrons exist in prokaryotes, archea, single-celled eukaryotes and multi-celled eukaryotes.
- 30 8. Self-limiting connectrons occur in prokaryotes, archea, single-celled eukaryotes and multi-celled eukaryotes
9. Geneless connectrons exist in single-celled and multi-celled eukaryotes

10. One connectron controls many geneless connectrons
in single-celled and multi-celled eukaryotes

1. Connectrons occur in prokaryotes, arch a,
single-celled eukaryotes and multi-celled
eukaryotes.

5

Connectrons exist as tetradic relationships where the
sequence T1 is equivalent to the sequence C1 (written
T1=C1) and where the sequence T2 equals the sequence C2
(written T2=C2) where T1 and T2 are DNA sequences 20 or
10 more bases in length, where the C1 sequence is adjacent
to the C2 sequence, where the T1 and T2 sequences are on
the same chromosome, and where the C1/C2 sequences are on
the same chromosome as T1 and T2 or where the C1/C2
sequences are on a chromosome different from T1 and T2.
15 The connectron relationship has been found to exist in
prokaryotes, archea, single-celled eukaryotes and multi-
celled eukaryotes.

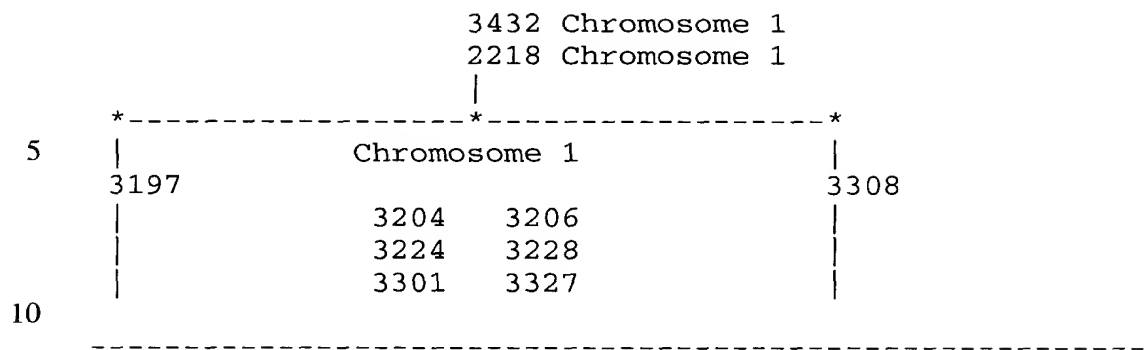
Example of a prokaryote connectron - E. coli

20

In this example the existence of the T1-T2 (3197-3308)
long loop is controlled by three C1/C2 short loops (3307,
3432 and 2218). The T1-T2 long loop controls the
expression of 64 genes on chromosome 1 in addition to six
25 C1/C2 (3204, 3206, 3223, 3228, 3301 and 3327) short
loops. The C1/C2 short loop 3327 lies outside the range
of the T1-T2 long loop (3197-3308) but this C1/C2 is
expressed as a 3'UTR to the gene hemG that is within the
range of the T1-T2 long loop.

30

3307 Chromosome 1



Connectron control elements for chromosome 1 of the E. coli genome

15 A double stranded DNA loop of length 93.542 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 3197. This T1 control element has the DNA sequence

20 Seq. Id. = 1 Position = 1 to 175

AAAAAATGCGCGGTCAGAAAATTATTTTAAATTTCTCTTGTCAGGCCGGAATAACT
CCCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACGCCGCCGGGTCAGCG
25 GGGTTCTCCTGAGAACTCCGGCAGAGAAAGCAAAAATAAATGCTTGACTCTGTAGCG
GGAA

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 3308. This T2
30 control element has the DNA sequence

Seq. Id. = 2 Position = 1 to 175

TAAATTTCTCTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTGACACGG
35 AACACGGCAAACACGCCGCCGGGTCAGCGGGTTCTCCTGAGAACTCCGGCAGAGA

AAGCAAAAATAAATGCTTGACTCTGTAGCGGGAAGGCGTATTATGCACACCCCGCGC
CGCT

This long T1/T2 double stranded DNA loop modulates the
5 expression of the following genes

	rrsC	gltU	rrlC	rrfC	aspT
	trpT	yifA	yifE	yifB	ilvL
	ilvG_1	ilvM	ilvE	ilvD	ilvA
10	ilvY	ilvC	ppiC	b3776	rep
	gppA	rhlB	trxA	rhoL	rho
	rfe	wzzE	wecB	rffH	wecD
	wecE	wzxE	yifM_2	wecG	yifK
	argX	hisR	leuT	proM	aslB
15	aslA	hemY	hemX	hemD	cyaA
	cyaY	b3808	dapF	uvrD	b3814
	corA	yigF	yigG	rarD	yigI
	pldA	recQ	yigJ	yigK	pldB
	yigL	yigM	metR	metE	ysgA
20	udp	yigN	ubiE	yigP	b3836
	yigU	yigW_1	rfaH	yigC	ubiB
	fadA	fadB	pepQ	trkH	hemG

This long T1/T2 double stranded DNA loop modulates the
25 expression of the following C1/C2 short loops

A C1/C2 short loop on chromosome 1 whose identifier is
3204 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop is
30 expressed as a RNA single strand that is 3'UTR to the
gene rrsC and has the DNA sequence

Seq. Id. = 3 Position = 1 to 186

GATGTGCCCAGATGGGATTAGCTAGTAGGTGGGGTAACGGCTCACCTAGGCGACGAT
CCCTAGCTGGTCTGAGAGGATGACCAGCCACACTGGAAGTACGACACGGTCCAGACT
CCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCA
5 TGCCGCGTGTATGAA

A C1/C2 short loop on chromosome 1 whose identifier is
3206 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop is
10 expressed as a RNA single strand that is 3'UTR to the
gene rrsC and has the DNA sequence

Seq. Id. = 4 Position = 1 to 186

15 GTCCCCTTCGTCTAGAGGCCAGGACACCGCCCTTTCACGGCGGTAACAGGGGTTCG
AATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTCACCTGCCTTAATAT
CTCAAAACTCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCA
AGCTGAAAATTGAAA

20 A C1/C2 short loop on chromosome 1 whose identifier is
3223 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop is
expressed as a RNA single strand that is 3'UTR to the
gene rrlC and has the DNA sequence

25 Seq. Id. = 5 Position = 1 to 186

GCTGAAGTAGGTCCCAAGGGTATGGCTGTTTCGCCATTTAAAGTGGTACGCGAGCTGG
GTTTAGAACGTCGTGAGACAGTTCGGTCCCTATCTGCCGTGGGCGCTGGAGAACTGA
30 GGGGGGCTGCTCCTAGTACGAGAGGACCGGAGTGGACGCATCACTGGTGTTCGGGTT
GTCATGCCAATGGCA

A C1/C2 short loop on chromosome 1 whose identifier is 3225 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene rrlC and has the DNA sequence

Seq. Id. = 6 Position = 1 to 144

AAACAGAATTTGCCTGGCGGCCGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAAC
10 TCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTA
GGGAAGTGCCAGGCATCAAATTAAGCAGTA

A C1/C2 short loop on chromosome 1 whose identifier is 3228 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene rrfC and has the DNA sequence

Seq. Id. = 7 Position = 1 to 112

20 GGTCATAAAACCGGTGGTTGTAAAAGAATTTCGGTGGAGCGGTAGTTCAGTCGGTTAG
AATACCTGCCTGTCACGCAGGGGGTTCGCGGGTTCGAGTCCCGTCCGTTCCGCCAC

A C1/C2 short loop on chromosome 1 whose identifier is 3301 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene ubiB and has the DNA sequence

30 Seq. Id. = 8 Position = 1 to 57

TTATCGTGCCTACAAATAGTCCGAACCGTAGGCCGGATAAGGCGTTTACGCCGCATC

A C1/C2 short loop on chromosome 1 whose identifier is 3307 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene fadA and has the DNA sequence

Seq. Id. = 9 Position = 1 to 56

TGCCGGATGCGGCGTAAACGCCTTATCCGGCCTACGGTTCGGACTATTTGTAGGCA

10

A C1/C2 short loop on chromosome 1 whose identifier is 3327 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene hemG and has the DNA sequence

Seq. Id. = 10 Position = 1 to 347

AAAAAATGCGCGGTCAGAAAATTATTTTAAATTTCTCTTGTCAGGCCGGAATAACT
20 CCCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACGCCCGCGGGTCAGCG
GGGTTCTCTCTGAGAACTCCGGCAGAGAAAGCAAAAATAAATGCTTGACTCTGTAGCG
GGAAGGCGTATTATG...CCCGTCACACCATGGGAGTGGGTGCAAAAGAAGTAGGT
AGCTTAACCTTCGGGAGGGCGCTTACCACTTTGTGATTCATGACTGGGGTGAAGTCG
TAACAAGGTAACCGTAGGGGAACCTGCGGTTGGATCACCTCCTTACCTTAAAGAAGC
25 GTTCTTTG

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

30 A C1/C2 short loop on chromosome 1 whose identifier is 3307 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA

single strand that is 3'UTR to the gene hemG and has the DNA sequence

Seq. Id. =11 Position = 1 to 347

5

AAAAAATGCGCGGTCAGAAAATTATTTTAAATTTTCCTCTTGTCAGGCCGGAATAACT
CCCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACGCCGCCGGGTCAGCG
GGGTTCTCCTGAGAACTCCGGCAGAGAAAGCAAAAATAAATGCTTGACTCTGTAGCG
GGAAGGCGTATTATG...CCCGTCACACCATGGGAGTGGGTTGCAAAAGAAGTAGGT
10 AGCTTAACCTTTCGGGAGGGCGCTTACCACTTTGTGATTCATGACTGGGGTGAAGTCG
TAACAAGGTAACCGTAGGGGAACCTGCGGTTGGATCACCTCCTTACCTTAAAGAAGC
GTTCTTTG

15 The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 11 Position = 1 to 175

AAAAAATGCGCGGTCAGAAAATTATTTTAAATTTTCCTCTTGTCAGGCCGGAATAACT
20 CCCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACGCCGCCGGGTCAGCG
GGGTTCTCCTGAGAACTCCGGCAGAGAAAGCAAAAATAAATGCTTGACTCTGTAGCG
GGAA

25 The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. = 11 Position = 28 to 202

TAAATTTTCCTCTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTGACACGG
30 AACAACGGCAAACACGCCGCCGGGTCAGCGGGGTTCTCCTGAGAACTCCGGCAGAGA
AAGCAAAAATAAATGCTTGACTCTGTAGCGGAAGGCGTATTATGCACACCCCGCGC
CGCT

A C1/C2 short loop on chromosome 1 whose identifier is 3432 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene btuB and has the
5 DNA sequence

Seq. Id. = 12 Position = 1 to 335

10 TGCGCGGTCAGAAAATTATTTTAAATTTCTCTTGTCAGGCCGGAATAACTCCCTAT
AATGCGCCACCACTGACACGGAACAACGGCAAACACGCCGCCGGGTCAGCGGGGTTC
TCCTGAGAACTCCGGCAGAGAAAGCAAAAATAAATGCTTGACTCTGTAGCGGGAAGG
CGTATTATGCACACC...ACACCATGGGAGTGGGTGCAAAAGAAGTAGGTAGCTTA
ACCTTCGGGAGGGCGCTTACCACTTTGTGATTCATGACTGGGGTGAAGTCGTAACAA
GGTAACCGTAGGGGAACCTGCGGTTGGATCACCTCCTTACCTTAAAGAAGCGT

15

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 12 Position = 1 to 169

20

TGCGCGGTCAGAAAATTATTTTAAATTTCTCTTGTCAGGCCGGAATAACTCCCTAT
AATGCGCCACCACTGACACGGAACAACGGCAAACACGCCGCCGGGTCAGCGGGGTTC
TCCTGAGAACTCCGGCAGAGAAAGCAAAAATAAATGCTTGACTCTGTAGCGGGAA

25 The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 12 Position = 22 to 196

30 TAAATTTCTCTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTGACACGG
AACAACGGCAAACACGCCGCCGGGTCAGCGGGGTTCCTGAGAACTCCGGCAGAGA
AAGCAAAAATAAATGCTTGACTCTGTAGCGGGAAGGCGTATTATGCACACCCCGCGC
CGCT

A C1/C2 short loop on chromosome 1 whose identifier is 2218 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene clpB and has the DNA sequence

Seq. Id. = 13 Position = 1 to 72

10 CTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTGACACGGAACAACGGCA
AACACGCCGCCGGGC

The match between the T1 sequence and the C1/C2 sequence is

15

Seq. Id. = 13 Position = 1 to 72

CTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTGACACGGAACAACGGCA
AACACGCCGCCGGGC

20

The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 13 Position = 1 to 71

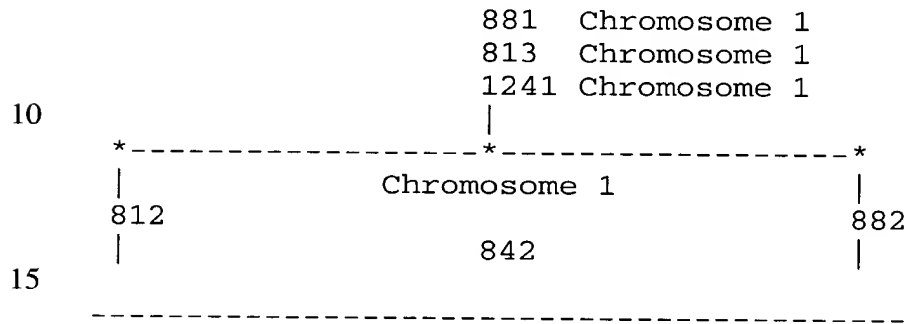
25

CTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTGACACGGAACAACGGCA
AACACGCCGCCGGG

30

Example of an archea connectron - *H. pylori*

In this example the existence of the T1-T2 (812-882) long loop is controlled by three C1/C2 short loops (881, 813 and 1214). The T1-T2 long loop controls the expression of 54 genes on chromosome 1 in addition to one C1/C2 (843) short loop.



Connectron control elements for chromosome 1 of *H. pylori* genome

A double stranded DNA loop of length 96.385 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 812. This T1 control element has the DNA sequence

Seq. Id. = 14 Position = 1 to 43

TTTTACTCATAGGGTTTTTATAGTTCCTAGCGGAAGTAAAGCA

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 882. This T2 control element has the DNA sequence

Seq. Id. = 15 Position = 1 to 43

TAGCGGAAGTAAAGCATTTCATCCCAAACACTAAAGATATTTGG

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

5	HP0999	HP1000	HP1001	HP1002	HP1003
	HP1005	HP1006	HP1008	HP1009	HPtRNA-Pro
	HP1010	HP1011	HP1013	HP1015	HP1017
	HP1018	HP1020	HP1021	HP1022	HP1023
	HP1024	HP1025	HP1027	HP1028	HP1030
10	HP1031	HP1033	HP1034	HP1038	HP1039
	HP1040	HP1041	HP1042	HP1043	HP1044
	HP1045	HP1046	HP1051	HP1052	HP1055
	HP1056	HP1058	HP1060	HP1065	HPtRNA-Ser
	HP1066	HP1067	HP1069	HP1070	HP1074
15	HP1075	HP1076	HP1077	HP1078	HP1079
	HP1080	HP1081	HP1083	HP1084	HP1085
	HP1088	HP1091	HP1092	HP1093	HP1094
	HP1095	HP1096			

20 This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

A C1/C2 short loop on chromosome 1 whose identifier is 813 controls the expression of the genes of one or more
 25 other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene HP0998 and has the DNA sequence

Seq. Id. = 16 Position = 1 to 70

30

TTTACTCATAGGGTTTTTATAGTTCCTAGCGGAATAAGCATTCATCCCAAACAC
 TAAAGATATTGG

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

5 A C1/C2 short loop on chromosome 1 whose identifier is 881 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene HP1096 and has the DNA sequence

10 Seq. Id. = 17 Position = 1 to 70

TTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCATTCATCCCAAACAC
TAAAGATATTTGG

15 The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 17 Position = 1 to 36

20 TTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCA

The match between the T2 sequence and the C1/C2 sequence is

25 Seq. Id. = 17 Position = 28 to 70

TAGCGGAACTAAAGCATTCATCCCAAACACTAAAGATATTTGG

30 The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 1 whose identifier is 813 controls the expression of the genes in this T1/T2

long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene HP0998 and has the DNA sequence

5 Seq. Id. = 18 Position = 1 to 70

TTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCATTCATCCCAAACAC
TAAAGATATTTGG

10 A C1/C2 short loop on chromosome 1 whose identifier is 881 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene HP1096 and has the DNA sequence

15

Seq. Id. = 19 Position = 1 to 70

TTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCATTCATCCCAAACAC
TAAAGATATTTGG

20

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 19 Position = 1 to 43

25

TTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCA

The match between the T2 sequence and the C1/C2 sequence is

30

Seq. Id. = 19 Position = 28 to 70

TAGCGGAACTAAAGCATTCATCCCAAACACTAAAGATATTTGG

A C1/C2 short loop on chromosome 1 whose identifier is 1241 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene HP1535 and has the DNA sequence

Seq. Id. = 20 Position = 1 to 56

10 TTTTACTCATAGGGTTTTTTATAGTTCCTAGCGGAACTAAAGCATTCATCCCAAACA

The match between the T1 sequence and the C1/C2 sequence is

15 Seq. Id. = 20 Position = 1 to 43

TTTTACTCATAGGGTTTTTTATAGTTCCTAGCGGAACTAAAGCA

The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 20 Position = 28 to 56

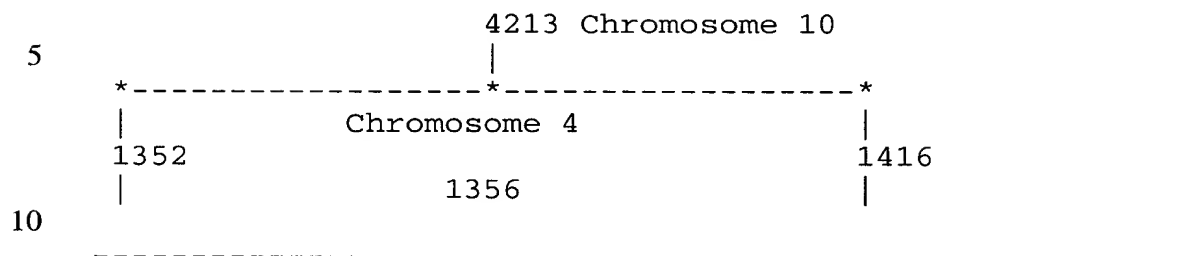
TAGCGGAACTAAAGCATTCATCCCAAACA

25

Example of single-celled connectron - *S. cerevisiae*

30 In this example the existence of the T1-T2 (1352-1416) long loop on chromosome 4 is controlled by one C1/C2 short loop (4213) on chromosome 10. The T1-T2 long loop

controls the expression of 34 genes on chromosome 4 in addition to one C1/C2 (1356) short loop.



Connectron control elements for chromosome 1 of *S. cerevisiae* genome

15

A double stranded DNA loop of length 68.908 kilo-bases on chromosome 4 is bounded on the left by a T1 sequence whose identifier is 1352. This T1 control element has the DNA sequence

20

Seq. Id. = 21 Position = 1 to 37

TTATGAGAAGCTGTCATCGAAGTTAGAGGAAGCTGAA

25

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 1416. This T2 control element has the DNA sequence

Seq. Id. = 22 Position = 1 to 362

30

ATTAGATCTATTACATTATGGGTGGTATGTTGGAATAAAAATCAACTATCATCTACT
 AACTAGTATTTACGTTACTAGTATATTATCATATACGGTGTAGAGAAGATGACGCAA
 TGATGAGAAATAGTCATCTAAATTAGTGGAAGCTGAAACGCAAGGATTGATAATGTA
 ATAGGATCAATGAATATTAACATATAAAACGATGATAATAATATTTATAGAATTGTG
 35 TAGAATTGCAGATTCCCTTTTATGGATTCCTAAATCCTTGAGGAGAACTTCTAGTAT

ATCTACATACCTAATATTATAGCCTTAATCACAATGGAATCCCAACAATTACATCAA
AATCCACATTCTCTACAGTA

This long T1/T2 double stranded DNA loop modulates the
5 expression of the following genes

	YDR170W-A	YDR171W	YDR172W	YDR173C	YDR174W
	YDR175C	YDR176W	YDR177W	YDR178W	YDR179C
	YDR179W-A	YDR180W	YDR181C	YDR182W	YDR183W
10	YDR184C	YDR185C	YDR186C	YDR187C	YDR188W
	YDR189W	YDR190C	YDR191W	YDR192C	YDR193W
	YDR194C	YDR195W	YDR196C	YDR197W	YDR198C
	YDR199W	YDR200C	YDR201W	YDR202C	YDR203W
	YDR204W	YDR205W	YDR206W	YDR207C	YDR208W
15	YDR209C	YDR210W			

This long T1/T2 double stranded DNA loop modulates the
expression of the following C1/C2 short loops

20 A C1/C2 short loop on chromosome 4 whose identifier is
1356 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop is
expressed as a RNA single strand that is 3'UTR to the
gene YDR170W-A and has the DNA sequence

25

Seq. Id. = 23 Position = 1 to 311

30 AATCACACTAATCATTCTGATGATGAACTCCCTGGACACCTCCTTCTCGATTCAGGA
GCATCACGAACCCTTATAAGATCTGCTCATCACATACTCAGCATCATCTAATCCT
GACATAAACGTAGTTGATGCTCAAAAAAGAAATATACCAATTAACGCTATTGGTGAC
CTACAATTTCACTTCCAGGACAACACCAAAACATCAATAAAGGTATTGCACACTCCT
AACATAGCCTATGACTTACTCAGTTTGAATGAATTGGCTGCAGTAGATATCACAGCA
TGCTTTACCAAAAACGTCTTAGAACG

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

5 A C1/C2 short loop on chromosome 10 whose identifier is 4213 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YJR029W and has the DNA sequence

10

Seq. Id. = 24 Position = 1 to 346

ATCTATTACATTATGGGTGGTATGTTGGAATAAAAATCCACTATCGTCTATCAACTA
ATAGTTATATTATCAATATATTATCATATACGGTGTTAAGATGATGACATAAGTTAT
15 GAGAAGCTGTCATCGAAGTTAGAGGAAGCTGAAACGCAAGGATTGATAATGTAATAG
GATCAATGAATATAAACATATAAACGGAATGAGGAATAATCGTAATATTAGTATGT
AGAAATATAGATTCCATTTTGAGGATTCCTATATCCTCGAGGAGAACTTCTAGTATA
TTCTGTATACCTAATATTATAGCCTTTATCAACAATGGAATCCCAACAATTATCTCA
ACAT

20

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 24 Position = 111 to 147

25

TTATGAGAAGCTGTCATCGAAGTTAGAGGAAGCTGAA

The match between the T2 sequence and the C1/C2 sequence is

30

Seq. Id. = 24 Position = 1 to 38

ATCTATTACATTATGGGTGGTATGTTGGAATAAAAATC

5

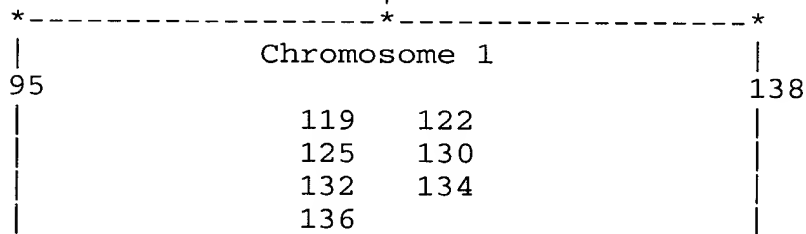
Example of a multi-celled connectron - *C. elegans*

In this example the existence of the T1-T2 (9-138) long
loop on chromosome 1 is controlled by three C1/C2 short
10 loops on chromosome 5 (21719, 21949 and 21655). The T1-
T2 long loop controls the expression of four genes on
chromosome 1 in addition to seven C1/C2 (119, 122, 125,
130, 132, 134 and 136) short loops.

15

21719 Chromosome 5
21949 Chromosome 5
21655 Chromosome 5

20



25

30

A double stranded DNA loop of length 41.978 kilo-bases on
chromosome 1 is bounded on the left by a T1 sequence
whose identifier is 95. This T1 control element has the
DNA sequence

Seq. Id. = 25 Position = 1 to 55

35

CAGCACGTTCTTAACCATGCAAAATCAGTTGAGAACTCTGCGTCTCTTCTCCCGC

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 138. This T2 control element has the DNA sequence

5 Seq. Id. = 26 Position = 1 to 36

ACTCTGCGTCTCTTCTCCCGCATTTTTGTAGATCA

10 This long T1/T2 double stranded DNA loop modulates the expression of the following genes

Y73A3A.1 Y73A3A.1 ZC123.3 ZC123.2

15 This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

20 A C1/C2 short loop on chromosome 1 whose identifier is 119 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene ZC123.3 and has the DNA sequence

Seq. Id. = 27 Position = 1 to 69

25 TTGAGAACTCTGCGTCTCAACTCCCGCATTTTTGTAGATCTACGTAGATCAAACCG
AAATGGGACACT

30 A C1/C2 short loop on chromosome 1 whose identifier is 122 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene ZC123.3 and has the DNA sequence

Seq. Id. = 28 Position = 1 to 89

GCACGGGGTTCTGGCCTTCCTCATTTGAATTTTTCGCGCTCCATTGACAATCGCCTGC
CGGACAACGCGTGGGAAAGTCGTGTACTCCAC

5

A C1/C2 short loop on chromosome 1 whose identifier is
125 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop is
expressed as a RNA single strand that is 3'UTR to the
10 gene ZC123.3 and has the DNA sequence

Seq. Id. = 29 Position = 1 to 89

ACGCGCCGTAAATCTACCCCAGATATGGCCGAGCCAAAATGGCCTAGTTCGGCAAAC
15 TCTTTCATTTCAATTTATGAGGGAAGCCAGAA

A C1/C2 short loop on chromosome 1 whose identifier is
130 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop is
20 expressed as a RNA single strand that is 3'UTR to the
gene ZC123.2 and has the DNA sequence

Seq. Id. = 30 Position = 1 to 121

25 CTCCCGCATTTTTTGTAGATCTACGTAGATCAAACCGAAATGAGGCACTTTCTGAAT
CCACGAGCTAGGCTTAAGCTTAGGCTTAAGCTTAGGCCTTTTCTCAGGCTTAGGCTT
AGGCTTA

A C1/C2 short loop on chromosome 1 whose identifier is
30 132 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop is
expressed as a RNA single strand that is 3'UTR to the
gene ZC123.2 and has the DNA sequence

Seq. Id. = 31 Position = 1 to 190

5 GCTTATGCTTGGGCTTAGGCTTAGGCGTAGGCTTAGGCTTAGGCTTAGGCTTATGCT
TAGACTTAGTCTCACTATCAGTCTTAGGCTTAGGCTTAGACTTAGGCTTAAGCTTAG
GCTTAAGCTTAGACTTAGGCTTAGGCTTAGGCTTAGGCTTAGGCTTAGGTTTGGGCT
TAGGCTTAGGCTTAACCTC

10 A C1/C2 short loop on chromosome 1 whose identifier is
134 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop is
expressed as a RNA single strand that is 3'UTR to the
gene ZC123.2 and has the DNA sequence

15 Seq. Id. = 32 Position = 1 to 133

20 TCTGCGTCTTTTCTCCCGCATTTTTTGTAGATCTACGTAGATCAAACCGAAATGAGG
CACTTTCTGAATCCACGAGCTAGGCTTAAGCTTAGGCTTAAGCTTAGGCCTTTTCTC
AGGCTTAGGCTTAGGCTTA

A C1/C2 short loop on chromosome 1 whose identifier is
136 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop is
expressed as a RNA single strand that is 3'UTR to the
25 gene ZC123.2 and has the DNA sequence

Seq. Id. = 33 Position = 1 to 190

30 GCTTATGCTTGGGCTTAGGCTTAGGCGTAGGCTTAGGCTTAGGCTTAGGCTTATGCT
TAGACTTAGTCTCACTATCAGTCTTAGGCTTAGGCTTAGACTTAGGCTTAAGCTTAG
GCTTAAGCTTAGACTTAGGCTTAGGCTTAGGCTTAGGCTTAGGCTTAGGTTTGGGCT
TAGGCTTAGGCTTAACCTC

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

5 A C1/C2 short loop on chromosome 5 whose identifier is 21719 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene C39F7.5 and has the DNA sequence

10 Seq. Id. = 34 Position = 1 to 65

ACGTTCTTAACCATGCAAAATCAGTTGAGAACTCTGCGTCTCTTCTCCCGCATTTTT
TGTAGATC

15 The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 34 Position = 1 to 51

20 ACGTTCTTAACCATGCAAAATCAGTTGAGAACTCTGCGTCTCTTCTCCCGC

The match between the T2 sequence and the C1/C2 sequence is

25 Seq. Id. = 34 Position = 31 to 65

ACTCTGCGTCTCTTCTCCCGCATTTTTTGTAGATC

30 A C1/C2 short loop on chromosome 5 whose identifier is 21949 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene F16B4.4 and has the DNA sequence

Seq. Id. = 35 Position = 1 to 95

ACCATGCAAAATCAGTTGAGAACTCTGCGTCTCTTCTCCCGCATTTTTTGTAGATCT
5 ACGTAGATCAAGCCGAAATGAGACACTCTGACACCACG

The match between the T1 sequence and the C1/C2 sequence
is

10 Seq. Id. = 35 Position = 1 to 42

ACCATGCAAAATCAGTTGAGAACTCTGCGTCTCTTCTCCCGC

The match between the T2 sequence and the C1/C2 sequence
15 is

Seq. Id. = 35 Position = 22 to 63

ACTCTGCGTCTCTTCTCCCGCATTTTTTGTAGATC

20

A C1/C2 short loop on chromosome 5 whose identifier is
21655 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene C39F7.3 and has
25 the DNA sequence

Seq. Id. = 36 Position = 1 to 61

AACCATGCAAAATCAGTTGAGAACTCTGCGTCTCTTCTCCCGCATTTTTTGTAGATC
30 TACG

The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 36 Position = 1 to 36

AACCATGCAAAATCAGTTGAGAACTCTGCGTCTCTTCTCCCGC

5

The match between the T2 sequence and the C1/C2 sequence
is

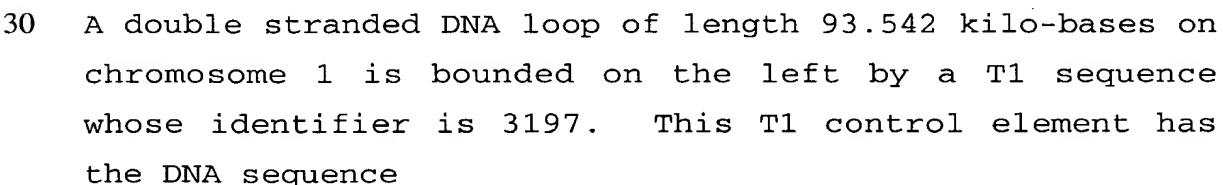
Seq. Id. = 36 Position = 23 to 57

10

ACTCTGCGTCTCTTCTCCCGCATTTTTGTAGATC

5 Many different C1/C2 short loops can control the
existence of one T1-T2 long loop. The C1/C2 short loops
can be on the same chromosome or on different chromosomes
from the T1-T2 long loop. This relationship is described
as "many-to-one". This relationship exists in
10 prokaryotes, archaea, single-celled eukaryotes and multi-
celled eukaryotes

15 In this example the existence of the T1-T2 (3197-3308)
long loop is controlled by three C1/C2 short loops (3307,
3432 and 2218).



- 68 -

AAAAAATGCGCGGTCAGAAAATTATTTTAAATTTCTCTTGTCAGGCCGGAATAACT
 CCCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACGCCGCCGGGTCAGCG
 GGGTTCTCCTGAGAACTCCGGCAGAGAAAGCAAAAATAAATGCTTGACTCTGTAGCG
 GGAA

5

This double stranded DNA loop is bounded on the right by
 a T2 control element whose identifier is 3308. This T2
 control element has the DNA sequence

10 Seq. Id. = 38 Position = 1 to 175

TAAATTTCTCTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTGACACGG
 AACAAACGGCAAACACGCCGCCGGGTCAGCGGGGTTCTCCTGAGAACTCCGGCAGAGA
 AAGCAAAAATAAATGCTTGACTCTGTAGCGGGAAGGCGTATTATGCACACCCCGCGC
 15 CGCT

This long T1/T2 double stranded DNA loop modulates the
 expression of the following genes

20	rrsC	gltU	rrlC	rrfC	aspT
	trpT	yifA	yifE	yifB	ilvL
	ilvG_1	ilvM	ilvE	ilvD	ilvA
	ilvY	ilvC	ppiC	b3776	rep
	gppA	rhlB	trxA	rhoL	rho
25	rfe	wzzE	wecB	rffH	wecD
	wecE	wzxE	yifM_2	wecG	yifK
	argX	hisR	leuT	proM	aslB
	aslA	hemY	hemX	hemD	cyaA
	cyaY	b3808	dapF	uvrD	b3814
30	corA	yigF	yigG	rarD	yigI
	pldA	recQ	yigJ	yigK	pldB
	yigL	yigM	metR	metE	ysgA
	udp	yigN	ubiE	yigP	b3836

yigU	yigW_1	rfaH	yigC	ubiB
fadA	fadB	pepQ	trkH	hemG

The expression of genes in this T1/T2 long loop is
5 controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 1 whose identifier is
3307 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
10 single strand that is 3'UTR to the gene hemG and has the
DNA sequence

Seq. Id. = 39 Position = 1 to 440

15 AAAAAATGCGCGGTCAGAAAATTATTTTAAATTTCTCTTGTTCAGGCCGGAATAACT
CCCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACGCCGCCGGGTCAGCG
GGGTTCTCTCTGAGAACTCCGGCAGAGAAAGCAAAAATAAATGCTTGACTCTGTAGCG
GGAAGGCGTATTATG...GGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAG
TAATCGTGGATCAGAATGCCACGGTGAATACGTTCCCGGGCCTTGTACACACCGCCC
20 GTCACACCATGGGAGTGGGTTGCAAAAGAAGTAGGTAGCTTAACCTTCGGGAGGGCG
CTTACCACTTTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAACCGTAGGGGA
ACCTGCGGTTGGATCACCTCCTTACCTTAAAGAAGCGTTCTTTG

The match between the T1 sequence and the C1/C2 sequence
25 is

Seq. Id. = 39 Position = 1 to 175

30 AAAAAATGCGCGGTCAGAAAATTATTTTAAATTTCTCTTGTTCAGGCCGGAATAACT
CCCTATAATGCGCCACCACTGACACGGAACAACGGCAAACACGCCGCCGGGTCAGCG
GGGTTCTCTCTGAGAACTCCGGCAGAGAAAGCAAAAATAAATGCTTGACTCTGTAGCG
GGAA

The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 39 Position = 28 to 192

5

TAAATTCCTCTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTGACACGG
AACAACGGCAAACACGCCGCCGGGTTCAGCGGGGTTCCTGAGAACTCCGGCAGAGA
AAGCAAAAATAAATGCTTGACTCTGTAGCGGGAAGGCGTATTATGCACACCCCGCGC
CGCT

10

A C1/C2 short loop on chromosome 1 whose identifier is 3432 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene btuB and has the

15 DNA sequence

Seq. Id. = 40 Position = 1 to 335

20 TGCGCGGTCAGAAAATTATTTTAAATTCCTCTTGTCAGGCCGGAATAACTCCCTAT
AATGCGCCACCACTGACACGGAACAACGGCAAACACGCCGCCGGGTTCAGCGGGGTTC
TCCTGAGAACTCCGGCAGAGAAAGCAAAAATAAATGCTTGACTCTGTAGCGGGAAGG
CGTATTATGCACACC...ACACCATGGGAGTGGGTTGCAAAGAAGTAGGTAGCTTA
ACCTTCGGGAGGGCGCTTACCACTTTGTGATTCATGACTGGGGTGAAGTCGTAACAA
GGTAACCGTAGGGGAACCTGCGGTTGGATCACCTCCTTACCTTAAAGAAGCGT

25

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 40 Position = 1 to 169

30

TGCGCGGTCAGAAAATTATTTTAAATTCCTCTTGTCAGGCCGGAATAACTCCCTAT
AATGCGCCACCACTGACACGGAACAACGGCAAACACGCCGCCGGGTTCAGCGGGGTTC
TCCTGAGAACTCCGGCAGAGAAAGCAAAAATAAATGCTTGACTCTGTAGCGGGAA

The match between the T2 sequence and the C1/C2 sequence is

5 Seq. Id. = 40 Position = 22 to 196

TAAATTTCTCTTGTCTAGGCCGGAATAACTCCCTATAATGCGCCACCACTGACACGG
AACACGGCAAACACGCCGCCGGGTCAGCGGGGTCTCTCTGAGAACTCCGGCAGAGA
AAGCAAAAATAAATGCTTGACTCTGTAGCGGGAAGGCGTATTATGCACACCCCGCGC
10 CGCT

A C1/C2 short loop on chromosome 1 whose identifier is 2218 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA
15 single strand that is 3'UTR to the gene clpB and has the DNA sequence

Seq. Id. = 41 Position = 1 to 72

20 CTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTGACACGGAACAACGGCA
AACACGCCGCCGGGC

The match between the T1 sequence and the C1/C2 sequence is

25

Seq. Id. = 41 Position = 1 to 72

CTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTGACACGGAACAACGGCA
AACACGCCGCCGGGC

30

The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 41 Position = 1 to 72

CTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTGACACGGAACAACGGCA
AACACGCCGCCGGGC

5

Example of a many-to-one connectron in archea - M.
jannaschii

10

In this example the existence of the T1-T2 (1630-1643)
long loop is controlled by four C1/C2 short loops (1629,
1642, 124 and 1533).

15

1629 Chromosome 1
1642 Chromosome 1
124 Chromosome 1
1533 Chromosome 1

20

----------*
| Chromosome 1 |
1630 1643

25

A double stranded DNA loop of length 4.998 kilo-bases on
chromosome 1 is bounded on the left by a T1 sequence
whose identifier is 1630. This T1 control element has
the DNA sequence

30

Seq. Id. = 42 Position = 1 to 175

35

TTATTAATTAGTTCAAAGGATTTTTATTTAATTTCTAAGGGTTTGCTGGTTTGATTA
TTTAGAATATTTGAGTTTATTGAATTATTCAGATTTTAAAAATTAAGATTAATTAG
GAAAGGAAATAAGATTTCTCTAACAGACAAGTTAAATTTTGGATTTAAAAAGATAA
AAAT

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 1643. This T2 control element has the DNA sequence

5

Seq. Id. = 43 Position = 1 to 175

TTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAATTA
TTCAGATTTTAAAAATTAGGATTAATTAGGCAAGTAAATAAAATTTCTCTAACAAA
10 TAAGTTAAATTTTGGATTAAAAAGATAAAAATACTCTGTTTTATTATGGAAAGAA
AGAT

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

15

MJ1597 MJ1598 MJ1599 MJ1600 MJ1601
MJ1602

20 The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 1 whose identifier is 1629 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA
25 single strand that is 3'UTR to the gene MJ1597 and has the DNA sequence

Seq. Id. = 44 Position = 1 to 139

30 ATATGTTTGAAATTTGAAAATAAGAGTATTTAGAAGTTATTAATTAGTTCAAAGGAT
TTTTATTTAATTTCTAAGGGTTGCTGGTTTGATTATTTAGAATATTTGAGTTTATT
GAATTATTCAGATTTTAAAAATTA

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 44 Position = 37 to 139

5

TTATTAATTAGTTCAAAGGATTTTATTTAATTTCTAAGGGTTTGCTGGTTTGATTA
TTTAGAATATTTGAGTTTATTGAATTATTCAGATTTTAAAAATTA

10 The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 44 Position = 81 to 139

15

GCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAATTATTCAGATTTTAAAAAT
TA

20 A C1/C2 short loop on chromosome 1 whose identifier is 1642 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene MJ1602 and has the DNA sequence

Seq. Id. = 45 Position = 1 to 177

25

ATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAAT
TATTCAGATTTTAAAAATTAGGATTAATTAGGCAAGTAAATAAAATTTCTCTAACA
AATAAGTTAAATTTTGGATTAAAAAGATAAAAAATACTCTGTTTTATTATGGAAAG
AAAGAT

30 The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 45 Position = 20 to 78

GCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAATTATTCAGATTTTAAAAAT
TA

5 The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. = 45 Position = 3 to 177

10 TTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAATTA
TTCAGATTTTAAAAATTAGGATTAATTAGGCAAGTAAATAAAATTTCTCTAACAAA
TAAGTTAAATTTTGGATTTAAAAAGATAAAAATACTCTGTTTTATTATGGAAAGAA
AGAT

15 A C1/C2 short loop on chromosome 1 whose identifier is
124 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene MJ0112 and has
the DNA sequence

20

Seq. Id. = 46 Position = 1 to 75

ATTTAATTTCTAAGGGTTTGCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAAT
TATTCAGATTTTAAAAAT

25

The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 46 Position = 1 to 75

30

ATTTAATTTCTAAGGGTTTGCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAAT
TATTCAGATTTTAAAAAT

The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 46 Position = 20 to 75

5

GCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAATTATTCAGATTTTAAAAAT

A C1/C2 short loop on chromosome 1 whose identifier is 1533 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene MJ1486 and has the DNA sequence

Seq. Id. = 47 Position = 1 to 58

15

TTTTTATTTAATTTCTAAGGGTTTGCTGGTTTGATTATTTAGAATATTTGAGTTTAT
T

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 47 Position = 1 to 58

TTTTTATTTAATTTCTAAGGGTTTGCTGGTTTGATTATTTAGAATATTTGAGTTTAT
25 T

The match between the T2 sequence and the C1/C2 sequence is

30 Seq. Id. = 47 Position = 25 to 58

GCTGGTTTGATTATTTAGAATATTTGAGTTTATT

Example of a many-to-one connectron in single-cell
eukaryotes - *S. cerevisiae*

5

In this example the existence of the T1-T2 (5515-5533)
long loop on chromosome 12 is controlled by seventeen
C1/C2 short loops (5516, 5532, 1939, 2323, 1942, 3286,
3649, 4764, 4751, 5536, 6102, 8023, 7356, 3293, 3291,
10 3289 and 146).

	5516 Chromosome 12
	5532 Chromosome 12
	1939 Chromosome 4
15	2323 Chromosome 5
	1942 Chromosome 5
	3286 Chromosome 7
	3649 Chromosome 8
	4764 Chromosome 12
20	4751 Chromosome 12
	5536 Chromosome 13
	6102 Chromosome 14
	8023 Chromosome 16
	7356 Chromosome 16
25	3293 Chromosome 8
	3291 Chromosome 8
	3289 Chromosome 8
	146 Chromosome 2
30	*-----*
	3197 3308
	Chromosome 12

35

A double stranded DNA loop of length 6.466 kilo-bases on
chromosome 12 is bounded on the left by a T1 sequence
whose identifier is 5515. This T1 control element has
the DNA sequence

40

Seq. Id. = 48 Position = 1 to 225

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
5 AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTG GTT

This double stranded DNA loop is bounded on the right by
a T2 control element whose identifier is 5533. This T2
control element has the DNA sequence

10

Seq. Id. = 49 Position = 1 to 225

ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTCTAGGGAATATGC
GTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTGTCACGGCAGTAGC
15 GAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAA
ATAAAGGTAGTAAGTAGCTTTTG GTTGAACATCCGGGTAAGAGACAACAGGGCT

This long T1/T2 double stranded DNA loop modulates the
expression of the following genes

20

YLR467W

This long T1/T2 double stranded DNA loop modulates the
expression of the following C1/C2 short loops

25

A C1/C2 short loop on chromosome 12 whose identifier is
5516 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop is
expressed as a RNA single strand that is 3'UTR to the
30 gene YLR464W and has the DNA sequence

Seq. Id. = 50 Position = 1 to 252

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAAC
5 ATCCGGGTAAGAGACAACAGGGCT

A C1/C2 short loop on chromosome 12 whose identifier is
5532 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop is
10 expressed as a RNA single strand that is 3'UTR to the
gene YLR467W and has the DNA sequence

Seq. Id. = 51 Position = 1 to 252

15 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAAC
ATCCGGGTAAGAGACAACAGGGCT

20

The expression of genes in this T1/T2 long loop is
controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 4 whose identifier is
25 1939 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene YDR545W and has
the DNA sequence

30 Seq. Id. = 52 Position = 1 to 222

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG

ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGG

5 The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 52 Position = 1 to 222

10 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGG

15 The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. = 52 Position = 28 to 222

20 ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTTCTAGGGAATATGC
GTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTGTCACGGCAGTAGC
GAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAA
ATAAAGGTAGTAAGTAGCTTTTGG

25 A C1/C2 short loop on chromosome 5 whose identifier is
2323 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene YER189W and has
the DNA sequence

30 Seq. Id. = 53 Position = 1 to 252

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG

ATTTAGTGTTTGTTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAAC
ATCCGGGTAAGAGACAACAGGGCT

- 5 The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 53 Position = 1 to 225

- 10 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTG

- 15 The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. = 53 Position = 28 to 252

- 20 ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTCTAGGGAATATGC
GTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTTCACGGCAGTAGC
GAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAA
ATAAAGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

- 25 A C1/C2 short loop on chromosome 5 whose identifier is
1942 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene YEL077C and has
the DNA sequence

30

Seq. Id. = 54 Position = 1 to 252

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAAC
5 ATCCGGGTAAGAGACAACAGGGCT

The match between the T1 sequence and the C1/C2 sequence
is

10 Seq. Id. = 54 Position = 1 to 225

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
15 AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTG

The match between the T2 sequence and the C1/C2 sequence
is

20 Seq. Id. = 54 Position = 28 to 252

ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTCTAGGGAATATGC
GTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTGTCACGGCAGTAGC
GAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAA
25 ATAAAGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

A C1/C2 short loop on chromosome 7 whose identifier is
3286 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
30 single strand that is 3'UTR to the gene YGR296W and has
the DNA sequence

Seq. Id. = 55 Position = 1 to 252

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
5 AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAAC
ATCCGGGTAAGAGACAACAGGGCT

The match between the T1 sequence and the C1/C2 sequence
is

10

Seq. Id. = 55 Position = 1 to 225

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
15 ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTG

The match between the T2 sequence and the C1/C2 sequence
is

20

Seq. Id. = 55 Position = 28 to 252

ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTCTAGGGAATATGC
GTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTGTCACGGCAGTAGC
25 GAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAA
ATAAAGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

A C1/C2 short loop on chromosome 8 whose identifier is
3649 controls the expression of the genes in this T1/T2
30 long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene YHR219W and has
the DNA sequence

Seq. Id. = 56 Position = 1 to 252

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
5 ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAAC
ATCCGGGTAAGAGACAACAGGGCT

The match between the T1 sequence and the C1/C2 sequence
10 is

Seq. Id. = 56 Position = 1 to 225

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
15 AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTG

The match between the T2 sequence and the C1/C2 sequence
20 is

Seq. Id. = 56 Position = 28 to 252

ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTCTAGGGAATATGC
25 GTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTGTCACGGCAGTAGC
GAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAA
ATAAAGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

A C1/C2 short loop on chromosome 12 whose identifier is
30 4764 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene YLL066C and has
the DNA sequence

Seq. Id. = 57 Position = 1 to 252

5 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAAC
ATCCGGGTAAGAGACAACAGGGCT

10 The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 57 Position = 1 to 225

15 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTG

20 The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. = 57 Position = 28 to 252

25 ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTCTAGGGAATATGC
GTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTGTCACGGCAGTAGC
GAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAA
ATAAAGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

30 A C1/C2 short loop on chromosome 12 whose identifier is
4751 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA

single strand that is 3'UTR to the gene YLL067C and has the DNA sequence

Seq. Id. = 58 Position = 1 to 252

5

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAAC

10 ATCCGGGTAAGAGACAACAGGGCT

The match between the T1 sequence and the C1/C2 sequence is

15 Seq. Id. = 58 Position = 1 to 225

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
20 AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTG

The match between the T2 sequence and the C1/C2 sequence is

25 Seq. Id. = 58 Position = 28 to 252

ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTCTAGGGAATATGC
GTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTGTCACGGCAGTAGC
GAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAA
30 ATAAAGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

A C1/C2 short loop on chromosome 13 whose identifier is 5536 controls the expression of the genes in this T1/T2

long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YML133C and has the DNA sequence

5 Seq. Id. = 59 Position = 1 to 252

```
AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
10 AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAAC
ATCCGGGTAAGAGACAACAGGGCT
```

The match between the T1 sequence and the C1/C2 sequence is

15

Seq. Id. = 59 Position = 1 to 252

```
AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
20 ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTG
```

The match between the T2 sequence and the C1/C2 sequence is

25

Seq. Id. = 59 Position = 28 to 252

```
TATAGTATATTGTAAGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGT
ATTTCACTGTTTTGATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAA
30 AGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTA
GCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT
```

A C1/C2 short loop on chromosome 14 whose identifier is 6102 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YNL339C and has
5 the DNA sequence

Seq. Id. = 60 Position = 1 to 252

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
10 AGAAATTTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAAC
ATCCGGGTAAGAGACAACAGGGCT

15 The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 60 Position = 1 to 225

20 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTG

25 The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. = 60 Position = 28 to 252

30 ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTTCTAGGGAATATGC
GTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTGTCACGGCAGTAGC
GAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAA
ATAAAGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

A C1/C2 short loop on chromosome 16 whose identifier is 8023 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YPR204W and has the DNA sequence

Seq. Id. = 61 Position = 1 to 252

10 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAAC
ATCCGGGTAAGAGACAACAGGGCT

15

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 61 Position = 1 to 252

20

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTG

25

The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 61 Position = 28 to 252

30

ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTCTAGGGAATATGC
GTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTGTCACGGCAGTAGC

GAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAA
ATAAAGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

A C1/C2 short loop on chromosome 16 whose identifier is
5 7356 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene YPL283C and has
the DNA sequence

10 Seq. Id. = 62 Position = 1 to 252

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
15 AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAAC
ATCCGGGTAAGAGACAACAGGGCT

The match between the T1 sequence and the C1/C2 sequence
is

20

Seq. Id. = 62 Position = 1 to 225

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
25 ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTG

The match between the T2 sequence and the C1/C2 sequence
is

30

Seq. Id. = 62 Position = 28 to 252

ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTCTAGGGAATATGC
GTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTGTCACGGCAGTAGC
GAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAA
ATAAAGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

5

A C1/C2 short loop on chromosome 8 whose identifier is
3293 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene YHL050C and has
the DNA sequence

10

Seq. Id. = 63 Position = 1 to 89

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTT

15

The match between the T1 sequence and the C1/C2 sequence
is

20 Seq. Id. = 63 Position = 1 to 89

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTT

25 The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. = 63 Position = 28 to 89

30 ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTTCTAGGGAATATGC
GTTTT

A C1/C2 short loop on chromosome 8 whose identifier is 3291 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YHL050C and has
5 the DNA sequence

Seq. Id. = 64 Position = 1 to 87

ATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTTCACGGCAGTAGCGAGAGA
10 CAAGTGGGAAAGAGTAGGATAAAAAGACAA

The match between the T1 sequence and the C1/C2 sequence is

15 Seq. Id. = 64 Position = 1 to 87

ATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTTCACGGCAGTAGCGAGAGA
CAAGTGGGAAAGAGTAGGATAAAAAGACAA

20 The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 64 Position = 1 to 87

25 ATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTTCACGGCAGTAGCGAGAGA
CAAGTGGGAAAGAGTAGGATAAAAAGACAA

A C1/C2 short loop on chromosome 2 whose identifier is 145 controls the expression of the genes in this T1/T2
30 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YBL113C and has the DNA sequence

Seq. Id. = 65 Position = 1 to 73

CTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAACATCCGGG
TAAGAGACAACAGGCT

5

The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 65 Position = 1 to 47

10

CTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTG

The match between the T2 sequence and the C1/C2 sequence
is

15

Seq. Id. = 65 Position = 1 to 73

CTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAACATCCGGG
TAAGAGACAACAGGCT

20

A C1/C2 short loop on chromosome 8 whose identifier is
3289 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene YHL050C and has
the DNA sequence

25

Seq. Id. = 66 Position = 1 to 73

CTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAACATCCGGG
TAAGAGACAACAGGCT

30

The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 66 Position = 1 to 47

CTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTG

5

The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 66 Position = 1 to 73

10

CTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAACATCCGGG
TAAGAGACAACAGGCT

15 A C1/C2 short loop on chromosome 2 whose identifier is 146 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YBL113C and has the DNA sequence

20 Seq. Id. = 67 Position = 1 to 62

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAA

25 The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 67 Position = 1 to 62

30 AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAA

The match between the T2 sequence and the C1/C2 sequence is

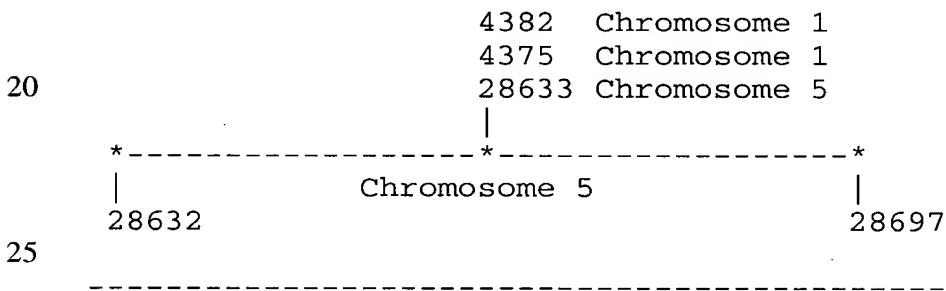
Seq. Id. = 67 Position = 28 to 62

5

ATTATGTATTGTGTAGTATAGTATATTGTAAGAAA

10 Example of a many-to-one connectron in multi-cell eukaryotes - C. elegans

In this example the existence of the T1-T2 (3197-3308) long loop on chromosome 5 is controlled by three C1/C2
15 short loops (4382, 4375 and 28633).



A double stranded DNA loop of length 58.451 kilo-bases on chromosome 5 is bounded on the left by a T1 sequence
30 whose identifier is 28632. This T1 control element has the DNA sequence

Seq. Id. = 68 Position = 1 to 86

35 GCAAAAATTGACTGAAAATTTGAATTTCCCGCAAAAATTGACTGAAAATTTGAATT
TCCCGCCAAAATTGACTGAAAATTTGAA

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 28697. This T2 control element has the DNA sequence

5 Seq. Id. = 69 Position = 1 to 160

CAAAAAATTGACTGAAAATTTGAATTTCCCTCCAAAAATTGACTGAAAATTTGAATT
TCCCGCCAAAAATTGACTGAAAATTTGAATATCCCGCCAAAAATTGACTGAAAATTT
GAATTTCCCGCCGAAAATTAAATGAAAAATGGAATTTCTCGCCGAA

10

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

M162.8 M162.4 M162.3 M162.6 M162.2
15 M162.1 M162.7

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

20 A C1/C2 short loop on chromosome 1 whose identifier is 4382 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene Y43F8B.10 and has the DNA sequence

25

Seq. Id. = 70 Position = 1 to 319

ATTATAGAAAATTTAAATTTCCCTCCAAAAATTGACTGAAAATTTGAATTTCCCTC
CAAAAAATTGACTGAAAATTTGAATTTCCCGCCAAAAATTGACTGAAAATTTGAATAT
30 CCCGCCAAAAATTGACTGAAAATTTGAATTTCCCGCCGAAAATTAAATGAAAAATGG
AATTTCTCGCCGAAAATTCAGTAAAAATTTGAATTTCTGCCAAAAATTGACTGAA
AATTTGAATTTCTTGCCAAAAAGTGACTGGGAATTTGAATTTCCCTCCAAAAATTG
ACTGAAATTTTGAATTTCCCGCTAAAAGTTGACT

The match between the T1 sequence and the C1/C2 sequence is

5 Seq. Id. = 70 Position = 58 to 88

CAAAAATTGACTGAAAATTTGAATTTCCCGC

10 The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 70 Position = 26 to 185

15 CAAAAAATTGACTGAAAATTTGAATTTCCCTCCAAAAATTGACTGAAAATTTGAATT
TCCCGCCAAAAATTGACTGAAAATTTGAATATCCCGCCAAAAATTGACTGAAAATTT
GAATTTCCCGCCGAAAATTAAATGAAAAATGGAATTTCTCGCCGAA

20 A C1/C2 short loop on chromosome 1 whose identifier is 4375 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene Y43F8B.10 and has the DNA sequence

25 Seq. Id. = 71 Position = 1 to 319

ATTATAGAAAATTTAAATTTCCCTCCAAAAAATTGACTGAAAATTTGAATTTCCCTC
CAAAAATTGACTGAAAATTTGAATTTCCCGCCAAAAATTGACTGAAAATTTGAATAT
CCCGCCAAAAATTGACTGAAAATTTGAATTTCCCGCCGAAAATTAAATGAAAAATGG
AATTTCTCGCCGAAAATTCAGTAAAAATTTGAATTTCCCTGCCAAAAATTGACTGAA
30 AATTTGAATTTCTTGCCAAAAAGTGACTGGGAATTTGAATTTCCCTCCAAAAATTG
ACTGAAATTTTGAATTTCCCGCTAAAAGTTGACT

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 71 Position = 58 to 88

5

CAAAAATTGACTGAAAATTTGAATTTCCCGC

The match between the T2 sequence and the C1/C2 sequence is

10

Seq. Id. = 71 Position = 58 to 217

CAAAAAATTGACTGAAAATTTGAATTTCCCTCCAAAAATTGACTGAAAATTTGAATT
TCCCGCCAAAAATTGACTGAAAATTTGAATATCCCGCCAAAAATTGACTGAAAATTT
15 GAATTTCCCGCCGAAAATTAAATGAAAAATGGAATTTCTCGCCGAA

A C1/C2 short loop on chromosome 5 whose identifier is 28633 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene M162.5 and has the DNA sequence

20

Seq. Id. = 72 Position = 1 to 85

25 CAAAAATTGACTGAAAATTTGAATTTCCCGCAAAAAATTGACTGAAAATTTGAATTT
CCCGCCAAAAATTGACTGAAAATTTGAA

Seq. Id. = 72 Position = 1 to 85

30 The match between the T1 sequence and the C1/C2 sequence is

CAAAAATTGACTGAAAATTTGAATTTCCCGCAAAAATTGACTGAAAATTTGAATTT
CCCGCCAAAATTGACTGAAAATTTGAA

5 The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. = 72 Position = 31 to 60

10 CAAAAATTGACTGAAAATTTGAATTTCCC

3. One connectron controls the expression of many sets of genes in prokaryotes, archea, single-celled eukaryotes and multi-celled eukaryotes.

5

One C1/C2 short loop can control the existence of a many T1-T2 long loops. The C1/C2 short loop can be on the same chromosome or on different chromosomes from the T1-T2 long loops. This relationship is described as "one-to-many". This relationship exists in prokaryotes, archea, single-celled eukaryotes and multi-celled eukaryotes.

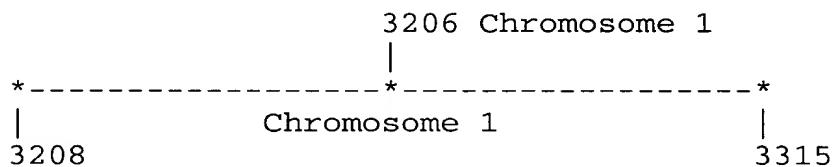
10

Example of a one-to-many connectron in prokaryotes - E. coli

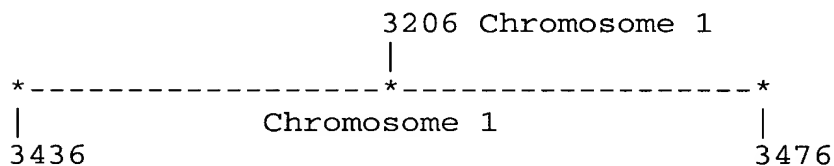
15

In this example the existence of T1-T2 (3208-3315, 3436-3476, 3439-3478 and 3441-3479) long loops are controlled by one C1/C2 short loop (3206).

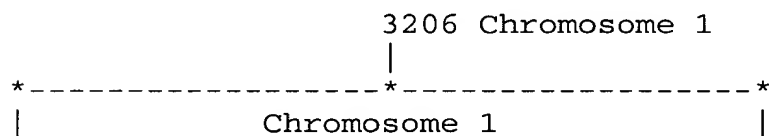
20



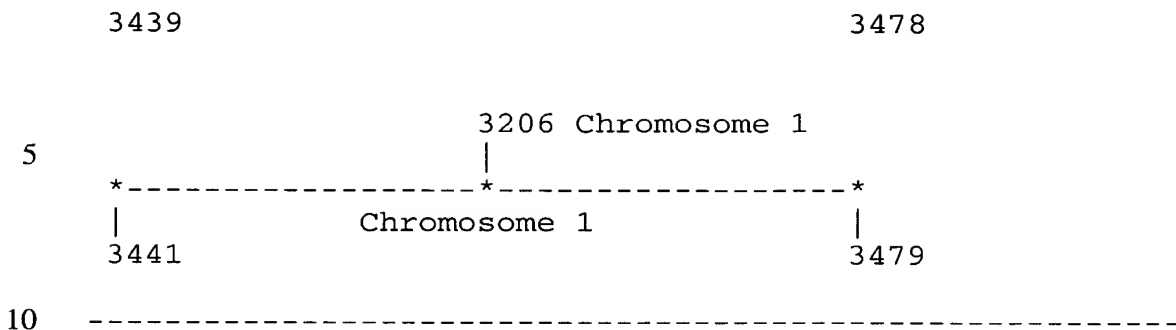
25



30



35



A double stranded DNA loop of length 93.377 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 3208. This T1 control element has the DNA sequence

Seq. Id. = 73 Position = 1 to 340

ACTCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCAAGCTGA
 AAATTGAAACACTGAACAACGAAAGTTGTTCGTGAGTCTCTCAAATTTTCGCAACAC
 GATGATGAATCGAAAGAAACATCTTCGGGTTGTGAGGTTAAGCGACTAAGCGTACAC
 GGTGGATGCCCTGGC...AGTGTGTTTCGACACACTATCATTAAGTGAATCCATAGG
 TTAATGAGGCGAACCGGGGGAAGTGAACATCTAAGTACCCCGAGGAAAAGAAATCA
 ACCGAGATTCCCCCAGTAGCGGCGAGCGAACGGGGAGCAGCCCAGAGCCTGAATCAG
 T

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 3315. This T2 control element has the DNA sequence

Seq. Id. = 74 Position = 1 to 330

TTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAACACTGAACAACGAAAGTT
 GTTCGTGAGTCTCTCAAATTTTCGCAACTCTGAAGTGAACATCTTCGGGTTGTGAG
 GTTAAGCGACTAAGCGTACACGGTGGATGCCCTGGCAGTCAGAGGCGATGAAGGACG
 TGCTAATCTGCGATA...GGTTAATGAGGCGAACCGGGGGAAGTGAACATCTAAGT

ACCCCGAGGAAAAGAAATCAACCGAGATTCCCCCAGTAGCGGCGAGCGAACGGGGAG
CAGCCCAGAGCCTGAATCAGTGTGTGTGTAGTGGAAGCGTCTGGAAA

This long T1/T2 double stranded DNA loop modulates the
5 expression of the following genes

	rrlC	rrfC	aspT	trpT	yifA
	yifE	yifB	ilvL	ilvG_1	ilvM
	ilvE	ilvD	ilvA	ilvY	ilvC
10	ppiC	b3776	rep	gppA	rhlB
	trxA	rhoL	rho	rfe	wzzE
	wecB	rffH	wecD	wecE	wzxE
	yifM_2	wecG	yifK	argX	hisR
	leuT	proM	aslB	aslA	hemY
15	hemX	hemD	cyaA	cyaY	b3808
	dapF	uvrD	b3814	corA	yigF
	yigG	rarD	yigI	pldA	recQ
	yigJ	yigK	pldB	yigL	yigM
	metR	metE	ysgA	udp	yigN
20	ubiE	yigP	b3836	yigU	yigW_1
	rfaH	yigC	ubiB	fadA	fadB
	pepQ	trkH	hemG	rrsA	ileT

The expression of genes in this T1/T2 long loop is
25 controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 1 whose identifier is
3206 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
30 single strand that is 3'UTR to the gene rrsC and has the
DNA sequence

Seq. Id. = 75 Position = 1 to 367

GTCCCCTTCGTCTAGAGGCCAGGACACCGCCCTTTCACGGCGGTAACAGGGGTTCG
 AATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTCACCTGCCTTAATAT
 CTCAAACTCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCA
 5 AGCTGAAAATTGAAA...ACCGGCGATTTCCGAATGGGGAAACCCAGTGTGTTTCGA
 CACACTATCATTAAGTGAATCCATAGGTTAATGAGGCGAACCGGGGGAAGTGAACAA
 TCTAAGTACCCCGAGGAAAAGAAATCAACCGAGATTCCCCCAGTAGCGGCGAGCGAA
 CGGGGAGCAGCCCAGAGCCTGAATCAGT

10 The match between the T1 sequence and the C1/C2 sequence
 is

Seq. Id. = 75 Position = 121 to 367

15 ACTCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCAAGCTGA
 AAATTGAAACACTGAACAACGAAAGTTGTTTCGTGAGTCTCTCAAATTTTCGCAACAC
 GATGATGAATCGAAAGAAACATCTTCGGGTTGTGAGGTTAAGCGACTAAGCGTACAC
 GGTGGATGCCCTGGC...AGTGTGTTTCGACACACTATCATTAAGTGAATCCATAGG
 TTAATGAGGCGAACCGGGGGAAGTGAACATCTAAGTACCCCGAGGAAAAGAAATCA
 20 ACCGAGATTCCCCCAGTAGCGGCGAGCGAACGGGGAGCAGCCCAGAGCCTGAATCAG
 T

The match between the T2 sequence and the C1/C2 sequence
 is

25 Seq. Id. = 75 Position = 148 to 232

TTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAACACTGAACAACGAAAGTT
 GTTCGTGAGTCTCTCAAATTTTCGCAAC

30 -----

A double stranded DNA loop of length 41.279 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 3436. This T1 control element has the DNA sequence

5

Seq. Id. = 76 Position = 1 to 113

ACGCAACGCGTGATAAGCAATTTTCGTGTCCCCTTCGTCTAGAGGCCAGGACACCG
CCCTTTCACGGCGGTAACAGGGGTTCGAATCCCCTAGGGGACGCCACTTGCTGGTT

10

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 3476. This T2 control element has the DNA sequence

15 Seq. Id. = 77 Position = 1 to 150

AGTGAAAAGCAAGGCGTCTTGCGAAGCAGACTGATACGTCCCCTTCGTCTAGAGGCC
CAGGACACCGCCCTTTCACGGCGGTAACAGGGGTTCGAATCCCCTAGGGGACGCCAC
TTGCTGGTTTGTGAGTGAAAGTCACCTGCCTTAATA

20

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

	gltT	rrlB	rrfB	murB	coaA
25	b3975	tyrU	thrT	tufB	secE
	nusG	rplK	rplA	rplJ	rplL
	rpoB	rpoC	htrC	thiH	thiF
	thiE	yjaE	yjaD	hemE	nfi
	yjaG	hupA	yjaH	yjaI	hydH
30	purD	purH			

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

A C1/C2 short loop on chromosome 1 whose identifier is 3206 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene rrsC and has the DNA sequence

Seq. Id. = 78 Position = 1 to 553

```
10 GTCCCCTTCGTCTAGAGGCCAGGACACCGCCCTTTCACGGCGGTAACAGGGGTTCG
AATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTCACCTGCCTTAATAT
CTCAAAACTCATCTTCGGGTGATGTTTGTGAGATATTTGCTCTTTAAAAATCTGGATCA
AGCTGAAAATTGAAACACTGAACAACGAAAGTTGTTTCGTGAGTCTCTCAAATTTTCG
CAACACGATGATGAATCGAAAGAAACATCTTCGGGTGAGGTTAAGCGACTAAGC
15 GTACACGGTGGATGCCCTGGCAGTCAGAGGCGATGAAGGACGTGCTAATCTGCGATA
AGCGTCGGTAAGGTGATATGAACCGTTATAACCGGCGATTTCCGAATGGGGAAACCC
AGTGTGTTTCGACACACTATCATTAAGTGAATCCATAGGTTAATGAGGCGAACCGGG
GGAAGTGAACATCTAAGTACCCCGAGGAAAAGAAATCAACCGAGATTCCCCCAGTA
GCGGCGAGCGAACGGGGAGCAGCCCAGAGCCTGAATCAGT
```

20

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 78 Position = 1 to 86

25

```
GTCCCCTTCGTCTAGAGGCCAGGACACCGCCCTTTCACGGCGGTAACAGGGGTTCG
AATCCCCTAGGGGACGCCACTTGCTGGTT
```

The match between the T2 sequence and the C1/C2 sequence is

30

Seq. Id. = 78 Position = 1 to 113

GTCCCCCTTCGTCTAGAGGCCAGGACACCGCCCTTTCACGGCGGTAACAGGGGTTTCG
AATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTCACCTGCCTTAATA

5

A double stranded DNA loop of length 41.336 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 3439. This T1 control element has the DNA sequence

10

Seq. Id. = 79 Position = 1 to 94

CCTTAATATCTCAAACTCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAA
TCTGGATCAAGCTGAAAATTGAAACACTGAACAACGA

15

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 3478. This T2 control element has the DNA sequence

20 Seq. Id. = 80 Position = 1 to 94

GTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAACAC
TGAACAACGAAAGTTGTTTCGTGAGTCTCTCAAATTTT

25 This long T1/T2 double stranded DNA loop modulates the expression of the following genes

	rrlB	rrfB	murB	coaA	b3975
	tyrU	thrT	tufB	secE	nusG
30	rplK	rplA	rplJ	rplL	rpoB
	rpoC	htrC	thiH	thiF	thiE
	yjaE	yjaD	hemE	nfi	yjaG

hupA yjaH yjaI hydH purD
 purH gltV

The expression of genes in this T1/T2 long loop is
 5 controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 1 whose identifier is
 3206 controls the expression of the genes in this T1/T2
 long loop. This C1/C2 short loop is expressed as a RNA
 10 single strand that is 3'UTR to the generrsC and has the
 DNA sequence

Seq. Id. = 81 Position = 1 to 367

15 GTCCCCTTCGTCTAGAGGCCAGGACACCGCCCTTTCACGGCGGTAACAGGGGTTTCG
 AATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTCACCTGCCTTAATAT
 CTCAAACTCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCA
 AGCTGAAAATTGAAA...ACCGGCGATTTCCGAATGGGGAAACCCAGTGTGTTTCGA
 CACACTATCATTAAGTGAATCCATAGGTTAATGAGGCGAACCAGGGGAACTGAAACA
 20 TCTAAGTACCCCGAGGAAAAGAAATCAACCGAGATTCCCCCAGTAGCGGCGAGCGAA
 CGGGGAGCAGCCAGAGCCTGAATCAGT

The match between the T1 sequence and the C1/C2 sequence
 is

25

Seq. Id. = 81 Position = 106 to 199

CCTTAATATCTCAAACTCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAA
 TCTGGATCAAGCTGAAAATTGAAACACTGAACAACGA

30

The match between the T2 sequence and the C1/C2 sequence
 is

Seq. Id. = 81 Position = 133 to 226

GTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAACAC
TGAACAACGAAAGTTGTTTCGTGAGTCTCTCAAATTTT

5

A double stranded DNA loop of length 38.285 kilo-bases on
chromosome 1 is bounded on the left by a T1 sequence
10 whose identifier is 3441. This T1 control element has
the DNA sequence

Seq. Id. = 82 Position = 1 to 355

15 AATTTTCGCAACACGATGATGAATCGAAAGAAACATCTTCGGGTTGTGAGGTTAAGC
GACTAAGCGTACACGGTGGATGCCCTGGCAGTCAGAGGCGATGAAGGACGTGCTAAT
CTGCGATAAGCGTCGGTAAGGTGATATGAACCGTTATAACCGGCGATTTCCGAATGG
GGAAACCCAGTGTGT...GATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAG
AACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCC
20 CACCTGACCCCATGCCGAACCTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGG
GGTCTCCCCATGCGAG

This double stranded DNA loop is bounded on the right by
a T2 control element whose identifier is 3479. This T2
25 control element has the DNA sequence

Seq. Id. = 83 Position = 1 to 356

AAGAAACATCTTCGGGTTGTGAGGTTAAGCGACTAAGCGTACACGGTGGATGCCCTG
30 GCAGTCAGAGGCGATGAAGGACGTGCTAATCTGCGATAAGCGTCGGTAAGGTGATAT
GAACCGTTATAACCGGCGATTTCCGAATGGGGAAACCCAGTGTGTTTCGACACACTA
TCATTAACTGAATCC...CAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAG
AATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACCTCAGAA

GTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAAC
TGCCAGGCATCAAATTA

5 This long T1/T2 double stranded DNA loop modulates the
expression of the following genes

rrlB	rrfB	murB	coaA	b3975
tyrU	thrT	tufB	secE	nusG
rplK	rplA	rplJ	rplL	rpoB
10 rpoC	htrC	thiH	thiF	thiE
yjaE	yjaD	hemE	nfi	yjaG
hupA	yjaH	yjaI	hydH	purD
purH	gltV			

15 The expression of genes in this T1/T2 long loop is
controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 1 whose identifier is
3206controls the expression of the genes in this T1/T2
20 long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene rrsC and has the
DNA sequence

Seq. Id. = 84 Position = 1 to 519

25

GTCCCCTTCGTCTAGAGGCCAGGACACCGCCCTTTCACGGCGGTAACAGGGGTTCG
AATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTCACCTGCCTTAATAT
CTCAAACTCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCA
AGCTGAAAATTGAAAAATTTTCGCAACACGATGATGAATCGAAAGAAACATCTTCGG
30 GTTGTGAGGTTAAGCGACTAAGCGTACACGGTGGATGCCCTGGCAGTCAGAGGCGAT
GAAGGACGTGCTAATCTGCGATAAGCGTCGGTAAGGTGATATGAACCGTTATAACCG
GCGATTTCCGAATGGGGAAACCCAGTGTGTTTCGACACACTATCATTAACCTGAATCC
ATAGGTTAATGAGGCGAACCGGGGGAACCTGAAACATCTAAGTACCCCGAGGAAAAGA

AATCAACCGAGATTCCCCCAGTAGCGGCGAGCGAACGGGGAGCAGCCCAGAGCCTGA
ATCAGT

5 The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 84 Position = 187 to 519

10 AATTTTCGCAACACGATGATGAATCGAAAGAAACATCTTCGGGTTGTGAGGTTAAGC
GACTAAGCGTACACGGTGGATGCCCTGGCAGTCAGAGGCGATGAAGGACGTGCTAAT
CTGCGATAAGCGTCGGTAAGGTGATATGAACCGTTATAACCGGCGATTTCCGAATGG
GGAAACCCAGTGTGTTTCGACACACTATCATTAAGTGAATCCATAGGTTAATGAGGC
GAACCGGGGGAAGTGAACATCTAAGTACCCCGAGGAAAAGAAATCAACCGAGATTC
CCCCAGTAGCGGCGAGCGAACGGGGAGCAGCCCAGAGCCTGAATCAGT

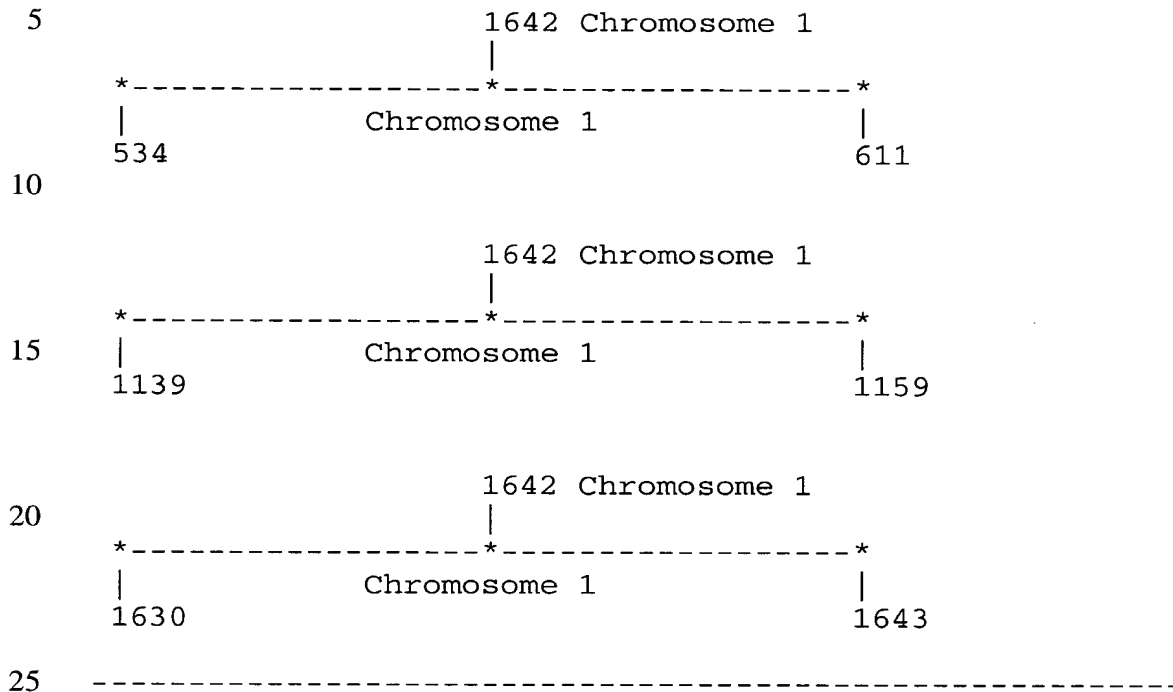
15 The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. = 84 Position = 214 to 519

20 AAGAAACATCTTCGGGTTGTGAGGTTAAGCGACTAAGCGTACACGGTGGATGCCCTG
GCAGTCAGAGGCGATGAAGGACGTGCTAATCTGCGATAAGCGTCGGTAAGGTGATAT
GAACCGTTATAACCGGCGATTTCCGAATGGGGAAACCCAGTGTGTTTCGACACACTA
TCATTAAGTGAATCCATAGGTTAATGAGGCGAACCGGGGGAAGTGAACATCTAAGT
25 ACCCCGAGGAAAAGAAATCAACCGAGATTCCCCCAGTAGCGGCGAGCGAACGGGGAG
CAGCCCAGAGCCTGAATCAGT

30 Example of a one-to-many connectron in archea - M.
jannaschii

In this example the existence of T1-T2 (534-611, 1139-1159, and 1630-1643) long loops are controlled by one C1/C2 short loop (1642).



A double stranded DNA loop of length 72.886 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 534. This T1 control element has the DNA sequence

Seq. Id. = 85 Position = 1 to 37

TAAGTAAATAAAATTTCTCTAACAAATAAGTTAAATT

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 611. This T2 control element has the DNA sequence

Seq. Id. = 86 Position = 1 to 59

TAAATAAAATTTCTCTAACAAATAAGTTAAATTTTGGATTAAAAAGATAAAAAATG
CT

5 This long T1/T2 double stranded DNA loop modulates the
expression of the following genes

	MJ0486	MJ0487	MJ0488	MJ0489	MJ0490
	MJ0492	MJ0493	MJ0494	MJ0495	MJ0496
10	MJ0497	MJ0499	MJ0500	MJ0501	MJ0502
	MJ0503	MJ0504	MJ0506	MJ0507	MJ0508
	MJ0509	MJ0510	MJ0511	MJ0512	MJ0513
	MJ0514	MJ0514	MJ0517	MJ0519	MJ0520
	MJ0521	MJ0522	MJ0523	MJ0525	MJ0526
15	MJ0526	MJ0529	MJ0530	MJ0531	MJ0532
	MJ0534	MJ0535	MJ0536	MJ0538	MJ0539
	MJ0540	MJ0541	MJ0542	MJ0543	MJ0544
	MJ0545	MJ0547	MJ0548	MJ0549	MJ0550
	MJ0552	MJ0553	MJ0554	MJ0555	MJ0556
20	MJ0558	MJ0559	MJ0560	MJ0561	MJ0562
	MJ0563	MJ0564			

The expression of genes in this T1/T2 long loop is
controlled by the following C1/C2 short loops.

25

A C1/C2 short loop on chromosome 1 whose identifier is
1642 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene MJ1602 and has

30

the DNA sequence

Seq. Id. = 87 Position = 1 to 177

ATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAAT
TATTCAGATTTTTTAAAAATTAGGATTAATTAGGCAAGTAAATAAAATTTCTCTAACA
AATAAGTTAAATTTTTGGATTTAAAAAGATAAAAATACTCTGTTTTATTATGGAAAG
AAAGAT

5

The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 87 Position = 92 to 127

10

AAGTAAATAAAATTTCTCTAACAAATAAGTTAAATT

The match between the T2 sequence and the C1/C2 sequence
is

15

Seq. Id. = 87 Position = 95 to 150

TAAATAAAATTTCTCTAACAAATAAGTTAAATTTTTGGATTTAAAAAGATAAAAAT

20

A double stranded DNA loop of length 14.509 kilo-bases on
chromosome 1 is bounded on the left by a T1 sequence
whose identifier is 1139. This T1 control element has
the DNA sequence

25

Seq. Id. = 88 Position = 1 to 78

ATTTATTAATTAGTTCAAAGGATTTTTATTTAATTTCTAAGGGTTAGCTGGTTTGAT
TGTTTAAATATTTGAGTTTA

30

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 1159. This T2 control element has the DNA sequence

5 Seq. Id. = 89 Position = 1 to 78

ATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAAT
TATTCAGATTTTTTAAAAATTA

10 This long T1/T2 double stranded DNA loop modulates the expression of the following genes

MJ1096	MJ1097	tRNA-Arg-3	MJ1098	MJ1099
MJ1100	MJ1101	MJ1102	MJ1103	MJ1104
15 MJ1105	MJ1106	MJ1107	MJ1108	

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

20 A C1/C2 short loop on chromosome 1 whose identifier is 1642 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene MJ1602 and has the DNA sequence

25

Seq. Id. = 90 Position = 1 to 177

ATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAAT
TATTCAGATTTTTTAAAAATTAGGATTAATTAGGCAAGTAAATAAAATTTCTCTAACA
30 AATAAGTTAAATTTTTGGATTTAAAAAGATAAAAATACTCTGTTTTATTATGGAAAG
AAAGAT

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 90 Position = 1 to 31

5

ATTTAATTTCTAAGGGTTAGCTGGTTTGATT

The match between the T2 sequence and the C1/C2 sequence is

10

Seq. Id. = 90 Position = 1 to 78

ATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAAT
TATTCAGATTTTAAAAATTA

15

A double stranded DNA loop of length 4.998 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 1630. This T1 control element has the DNA sequence

Seq. Id. = 91 Position = 1 to 175

25 TTATTAATTAGTTCAAAGGATTTTATTTAATTTCTAAGGGTTTGCTGGTTTGATTA
TTTAGAATATTTGAGTTTATTGAATTATTCAGATTTTAAAAATTAAGATTAATTAG
GAAAGGAAATAAGATTTCTCTAACAGACAAGTTAAATTTTGGATTTAAAAAGATAA
AAAT

30 This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 1643. This T2 control element has the DNA sequence

Seq. Id. = 92 Position = 1 to 175

TTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAATTA
TTCAGATTTTTTAAAAATTAGGATTAATTAGGCAAGTAAATAAAATTTCTCTAACAAA
5 TAAGTTAAATTTTTGGATTTAAAAAGATAAAAATACTCTGTTTTATTATGGAAAGAA
AGAT

This long T1/T2 double stranded DNA loop modulates the
expression of the following genes

10

MJ1597 MJ1598 MJ1599 MJ1600 MJ1601
MJ1602

The expression of genes in this T1/T2 long loop is
15 controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 1 whose identifier is
1642 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
20 single strand that is 3'UTR to the gene MJ1602 and has
the DNA sequence

Seq. Id. = 93 Position = 1 to 177

25 ATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAAT
TATTCAGATTTTTTAAAAATTAGGATTAATTAGGCAAGTAAATAAAATTTCTCTAACA
AATAAGTTAAATTTTTGGATTTAAAAAGATAAAAATACTCTGTTTTATTATGGAAAG
AAAGAT

30 The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 93 Position = 20 to 78

GCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAATTATTCAGATTTTAAAAAT
TA

5 The match between the T2 sequence and the C1/C2 sequence
is

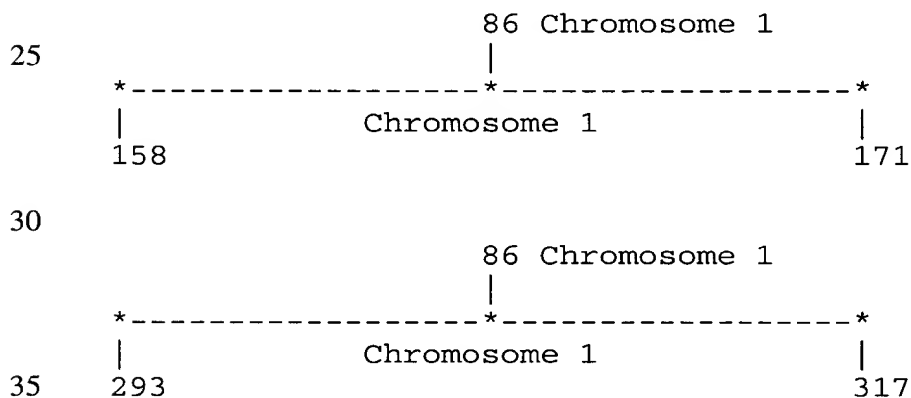
Seq. Id. = 93 Position = 3 to 177

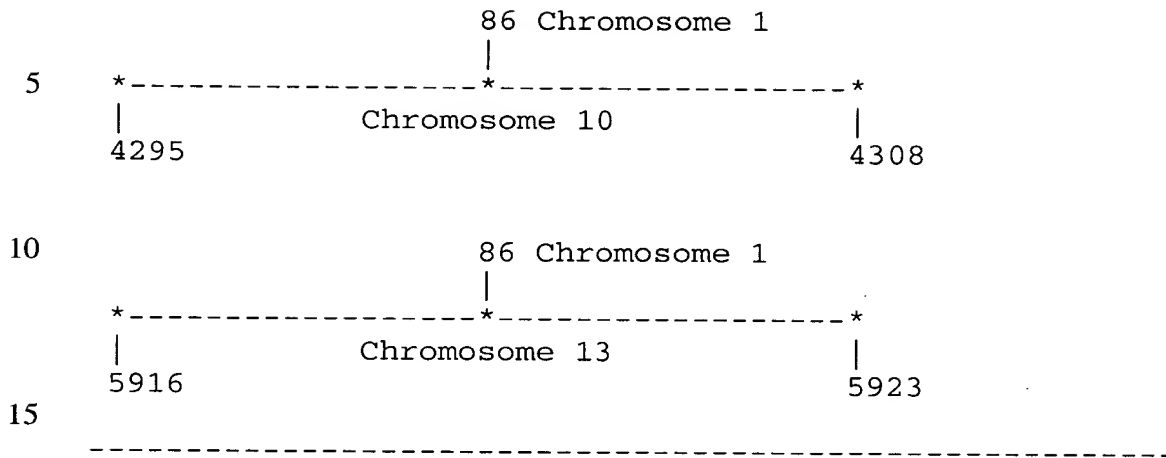
10 TTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAATTA
TTCAGATTTTAAAAATTAGGATTAATTAGGCAAGTAAATAAAATTTCTCTAACAAA
TAAGTTAAATTTTGGATTAAAAAGATAAAAATACTCTGTTTTATTATGGAAAGAA
AGAT

15 -----

Example of a one-to-many connectron in single-cell
eukaryotes - *S. cerevisiae*

20 In this example the existence of T1-T2 (158-171, 293-
317, 4295-4308 and 5916-5923) long loops are controlled
by one C1/C2 short loop (86).





A double stranded DNA loop of length 20.391 kilo-bases on chromosome 2 is bounded on the left by a T1 sequence whose identifier is 158. This T1 control element has the DNA sequence

Seq. Id. = 94 Position = 1 to 153

CCAATTGTTGGAATAAAAATCAACTATCATCTACTAACTAGTATTTACGTTACTAGT
ATATTATCATATACGGTGTTAGAAGATGACGCAAATGATGAGAAATAGTCATCTAAA
TTAGTGGAAGCTGAAACGCAAGGATTGATAATGTAATAG

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 171. This T2 control element has the DNA sequence

Seq. Id. = 95 Position = 1 to 192

ATAATTGTTGGAATAAAAATCAACTATCATCTACTAACTAGTATTTACGTTACTAGT
ATATTATCATATACGGTGTTAGAAGATGACACAAATGATGAGAAATAGTCATCTAAA
TTAGTGGAAGCTGAAACGCAAGGATTGATAATGTAATAGGATCAATGAATATTAACA
TATAAAATGATGATAATAATA

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

5 YBL107W-A TL(UAA)B1 YBL107C YBL106C YBL105C
YBL104C YBL103C YBL102W YBL101C

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

10

A C1/C2 short loop on chromosome 1 whose identifier is 86 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YAR009C and has the DNA
15 sequence

Seq. Id. = 96 Position = 1 to 362

ATCTATTACATTATGGGTGGTATGTTGGAATAGAAATCAACTATCATCTACTAACTA
20 GTATTTACATTACTAGTATATTATCATATACGGTGTTAGAAGATGACGCAAATGATG
AGAAATAGTCATCTAAATTAGTGGAAGCTGAAACGCAAGGATTGATAATGTAATAGG
ATCAATGAATATAAACATATAAAACGGAATGAGGAATAATCGTAATATTAGTATGTA
GAAATATAGATTCCATTTTGAGGATTCCTATATCCTCGAGGAGAACTTCTAGTATAT
TCTGTATACCTAATATTATAGCCTTTATCAACAATGGAATCCCAACAATTATCTCAA
25 CATTCACCCATTTCTCAGAA

The match between the T1 sequence and the C1/C2 sequence is

30 Seq. Id. = 96 Position = 34 to 65

AAATCAACTATCATCTACTAACTAGTATTTAC

The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 96 Position = 34 to 65

5

AAATCAACTATCATCTACTAACTAGTATTTAC

10 A double stranded DNA loop of length 38.470 kilo-bases on chromosome 2 is bounded on the left by a T1 sequence whose identifier is 293. This T1 control element has the DNA sequence

15 Seq. Id. = 97 Position = 1 to 258

GAATTGTTGGAATAAAAATCCACTATCGTCTATCAACTAATAGTTATATTATCAATA
TATTATCATATACGGTGTTAAGATGATGACATAAGTTATGAGAAGCTGTCATCGAAG
TTAGAGGAAGCTGAAGTGCAAGGATTGATAATGTAATAGGATAATGAAACATATAAA
20 ACGGAATGAGGAATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTTGAGG
ATTCCTATATCCTTGAGGAGAACTTCTAGT

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 317. This T2
25 control element has the DNA sequence

Seq. Id. = 98 Position = 1 to 77

AATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCTATATCCTCGAGGAG
30 AACTTCTAGTATATTCTGTA

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

	YBL005W-B	TS (AGA) B	YBL004W	YBL003C	YBL002W
	YBL001C	YBR001C	YBR002C	YBR003W	YBR004C
	YBR005W	YBR006W	YBR007C	YBR008C	YBR009C
5	YBR010W	YBR011C	YBR012C		

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

10 A C1/C2 short loop on chromosome 1 whose identifier is 86 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YAR009C and has the DNA sequence

15

Seq. Id. = 99 Position = 1 to 362

ATCTATTACATTATGGGTGGTATGTTGGAATAGAAATCAACTATCATCTACTAACTA
 GTATTTACATTACTAGTATATTATCATATACGGTGTTAGAAGATGACGCAAATGATG
 20 AGAAATAGTCATCTAAATTAGTGGAAAGCTGAAACGCAAGGATTGATAATGTAATAGG
 ATCAATGAATATAAACATATAAAACGGAATGAGGAATAATCGTAATATTAGTATGTA
 GAAATATAGATTCCATTTTGAGGATTCCTATATCCTCGAGGAGAACTTCTAGTATAT
 TCTGTATACCTAATATTATAGCCTTTATCAACAATGGAATCCCAACAATTATCTCAA
 CATTCACCCATTTCTCAGAA

25

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 99 Position = 181 to 264

30

AAACATATAAAACGGAATGAGGAATAATCGTAATATTAGTATGTAGAAATATAGATT
 CCATTTTGAGGATTCCTATATCCT

The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 99 Position = 215 to 291

5

AATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCTATATCCTCGAGGAG
AACTTCTAGTATATTCTGTA

10

A double stranded DNA loop of length 11.020 kilo-bases on chromosome 10 is bounded on the left by a T1 sequence whose identifier is 4295. This T1 control element has the DNA sequence

15

Seq. Id. = 100 Position = 1 to 145

AAACGCAAGGATTGATAATGTAATAGGATCAATGAATATAAACATATAAAACGGAAT
GAGGAATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCTA
20 TATCCTCGAGGAGAACTTCTAGTATATTCTG

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 4308. This T2 control element has the DNA sequence

25

Seq. Id. = 101 Position = 1 to 180

GGAAGCTGAAACGCAAGGATTGATAATGTAATAGGATCAATGAATATAAACATATAA
AACGGAATGAGGAATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTTGAG
30 GATTCCTATATCCTCGAGGAGAACTTCTAGTATATTCTGTATACCTAATATTATAGC
CTTTATCAA

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

YJR027W YJR029W

5

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

10 A C1/C2 short loop on chromosome 1 whose identifier is 87 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YAR009C and has the DNA sequence

15 Seq. Id. = 102 Position = 1 to 359

ATCTATTACATTATGGGTGGTATGTTGGAATAGAAATCAACTATCATCTACTAACTA
GTATTTACATTACTAGTATATTATCATATACGGTGTTAGAAGATGACGCAAATGATG
AGAAATAGTCATCTAAATTAGTGGAAGCTGAAACGCAAGGATTGATAATGTAATAGG
20 ATCAATGAATATAAACATATAAAACGGAATGAGGAATAATCGTAATATTAGTATGTA
GAAATATAGATTCCATTTTGAGGATTCCTATATCCTCGAGGAGAACTTCTAGTATAT
TCTGTATACCTAATATTATAGCCTTTATCAACAATGGAATCCCAACAATTATCTCAA
CATTCACCCATTTCTCA

25 -----

A double stranded DNA loop of length 5.462 kilo-bases on chromosome 13 is bounded on the left by a T1 sequence whose identifier is 5916. This T1 control element has
30 the DNA sequence

Seq. Id. = 103 Position = 1 to 146

AAGCTGAAGTGCAAGGATTGATAATGTAATAGGATAATGAAACATATAAAACGGAAT
GAGGAATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCTA
TATCCTCGAGGAGAACTTCTAGTATATTCTGTA

- 5 This double stranded DNA loop is bounded on the right by
a T2 control element whose identifier is 5923. This T2
control element has the DNA sequence

Seq. Id. = 104 Position = 1 to 146

10

104TAATAGGATAATGAAACATATAAAACGGAATGAGGAATAATCGTAATATTAGTA
TGTAGAAATATAGATTCCATTTTGAGGATTCCTATATCCTCGAGGAGAACTTCTAGT
ATATTCTGTATACCTAATATTATAGCCTTTATCAA

- 15 This long T1/T2 double stranded DNA loop modulates the
expression of the following genes

YML045W

- 20 The expression of genes in this T1/T2 long loop is
controlled by the following C1/C2 short loops.

- A C1/C2 short loop on chromosome 1 whose identifier is 87
controls the expression of the genes in this T1/T2 long
25 loop. This C1/C2 short loop is expressed as a RNA single
strand that is 3'UTR to the gene YAR009C and has the DNA
sequence

Seq. Id. = 105 Position = 1 to 359

30

ATCTATTACATTATGGGTGGTATGTTGGAATAGAAATCAACTATCATCTACTAACTA
GTATTTACATTACTAGTATATTATCATATACGGTGTTAGAAGATGACGCAAATGATG
AGAAATAGTCATCTAAATTAGTGGAAGCTGAAACGCAAGGATTGATAATGTAATAGG

ATCAATGAATATAAACATATAAAACGGAATGAGGAATAATCGTAATATTAGTATGTA
 GAAATATAGATTCCATTTTGAGGATTCCTATATCCTCGAGGAGAACTTCTAGTATAT
 TCTGTATACCTAATATTATAGCCTTTATCAACAATGGAATCCCAACAATTATCTCAA
 CATTCACCCATTTCTCA

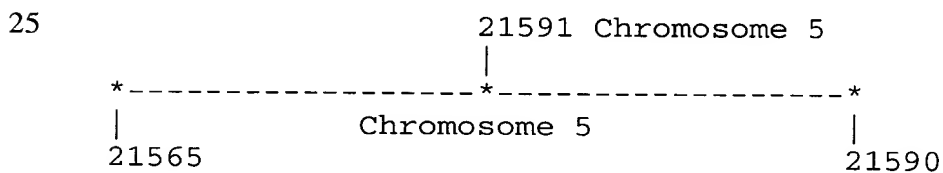
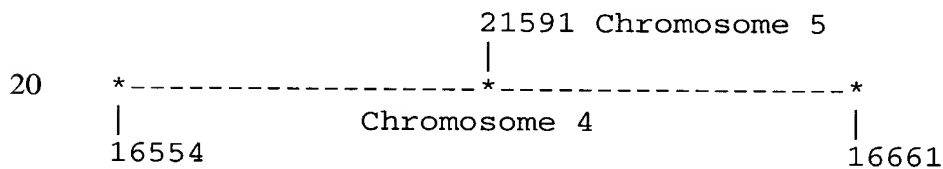
5

Example of a one-to-many connectron in multi-cell
 eukaryotes - C. elegans

10

In this example the existence of T1-T2 (16554-16661 and
 21565-21590) long loops are controlled by one C1/C2 short
 loop (21591).

15



30

A double stranded DNA loop of length 50.159 kilo-bases on
 chromosome 4 is bounded on the left by a T1 sequence
 whose identifier is 16554. This T1 control element has
 the DNA sequence

Seq. Id. = 106 Position = 1 to 143

TGCCTGAAAAAATTGGCTCCGAGTTAGGACACTTGGGGTGGTCAAAAAATTTTGTGA
CTATTGTCAAATGAAAGATCATAGTTGATAACATAAATTCCCAAAGTTTCATAAAAA
5 TCGATACGCAGCGAACAAAGTTATCAATT

This double stranded DNA loop is bounded on the right by
a T2 control element whose identifier is 16661. This T2
control element has the DNA sequence

10

Seq. Id. = 107 Position = 1 to 141

CACTTGGGGTGGTCAAAAAATTTTGTGATTATTGTCAAATGAAAGATCATGGTTGAT
AACATAAATTCCCAAAGTTTCATAAAAAATCGATACGCAGCGAACAAAGTTATGATTT
15 TTGACCCGGAAGTTATTTGGAGACCTA

This long T1/T2 double stranded DNA loop modulates the
expression of the following genes

20 C23H5.7 C23H5.8a C23H5.3 C23H5.2 C23H5.9
C23H5.1

The expression of genes in this T1/T2 long loop is
controlled by the following C1/C2 short loops.

25

A C1/C2 short loop on chromosome 5 whose identifier is
21591 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene F25A2.1 and has
30 the DNA sequence

Seq. Id. = 108 Position = 1 to 117

TATTGTCAAATGAAAGATCATGGTTGATAACATAAATTCCCACAATTCATAAAAAT
CGATACGCAGCGAACAAAGTTATGATTTTTGACCCGGAAC TTATTTGGAGACCTAAT
ATT

5 The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 108 Position = 46 to 85

10 TTTCATAAAAATCGATACGCAGCGAACAAAGTTAT

The match between the T2 sequence and the C1/C2 sequence
is

15 Seq. Id. = 108 Position = 1 to 42

TATTGTCAAATGAAAGATCATGGTTGATAACATAAATTCCCA

20

A double stranded DNA loop of length 18.142 kilo-bases on
chromosome 5 is bounded on the left by a T1 sequence
whose identifier is 21565. This T1 control element has
the DNA sequence

25

Seq. Id. = 109 Position = 1 to 72

CTCCGAGTTAGGACACTTGGGGTGGACAAAAAATTTGTGACTATTGTCAAATGAAA
GATCATGGTTGATAA

30

This double stranded DNA loop is bounded on the right by
a T2 control element whose identifier is 21590. This T2
control element has the DNA sequence

Seq. Id. = 110 Position = 1 to 115

TATTGTCAAATGAAAGATCATGGTTGATAACATAAATTCCCACAATTCATAAAAAAT
5 CGATACGCAGCGAACAAGTTATGATTTTTGACCCGGAACCTATTTGGAGACCTAAT
A

This long T1/T2 double stranded DNA loop modulates the
expression of the following genes

10

T21H3.2 T21H3.1 F25A2.1

The expression of genes in this T1/T2 long loop is
controlled by the following C1/C2 short loops.

15

A C1/C2 short loop on chromosome 5 whose identifier is
21591 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene F25A2.1 and has
20 the DNA sequence

Seq. Id. = 111 Position = 1 to 117

TATTGTCAAATGAAAGATCATGGTTGATAACATAAATTCCCACAATTCATAAAAAAT
25 CGATACGCAGCGAACAAGTTATGATTTTTGACCCGGAACCTATTTGGAGACCTAAT
ATT

The match between the T1 sequence and the C1/C2 sequence
is

30

Seq. Id. = 111 Position = 1 to 30

TATTGTCAAATGAAAGATCATGGTTGATAA

The match between the T2 sequence and the C1/C2 sequence
is

5 Seq. Id. = 111 Position = 1 to 115

TATTGTCAAATGAAAGATCATGGTTGATAACATAAATTCCCACAATTCATAAAAAT
CGATACGCAGCGAACAAAGTTATGATTTTTGACCCGGAAGTTATTTGGAGACCTAAT
A

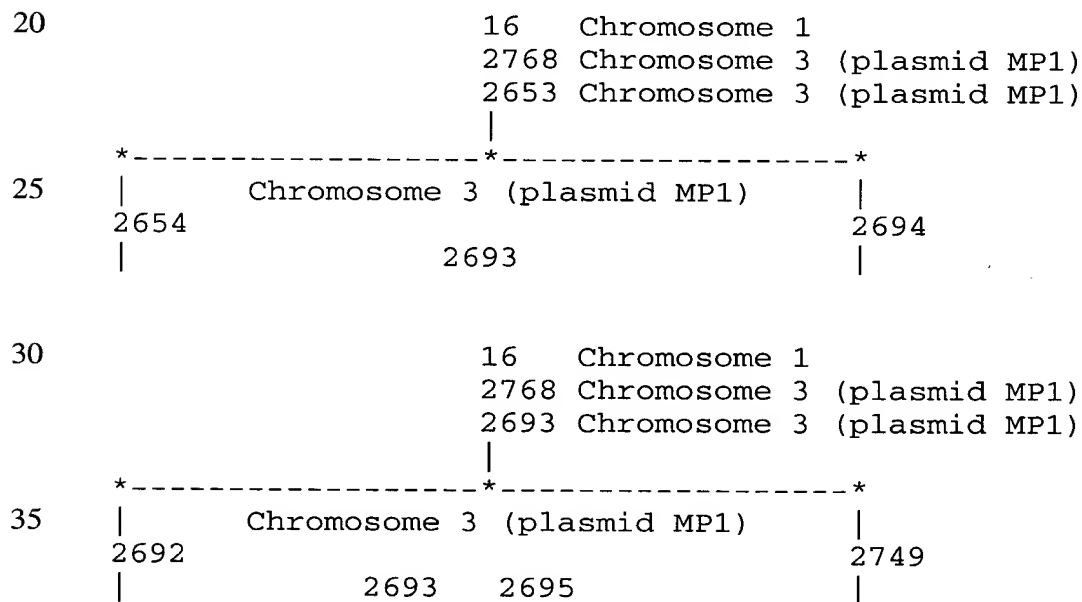
10

4. Connectrons occur between prokaryotes and their plasmids.

Connectron relationships exist between prokaryotes and their plasmids. These connectrons implement a control mechanism between the two genomes that makes it possible for them to form a symbiotic relationship. In the case of *D. radiodurans* the relationship is not symmetric. The *D. radiodurans* genome sends C1/C2 short loops to the MP1 plasmid.

Example of a prokaryote/plasmid connectron - *D. radiodurans*

In this example the existence of T1-T2 (2654-2694 and 2692-2749) long loops in chromosome 3 that is the plasmid MP1 are controlled by one C1/C2 short loop (16) in chromosome 1.



A double stranded DNA loop of length 46.903 kilo-bases on chromosome 3 (plasmid MP1) is bounded on the left by a T1 sequence whose identifier is 2654. This T1 control
5 element has the DNA sequence

Seq. Id. = 112 Position = 1 to 274

10 CAGCGTTTTTCTCGCTGTTCTGACGGCTGAACGCCCTGAATCTCTCCCGGTATGC
AGCCTGCTCGGAGAGTACGATTCGTTCGTGCTGCACCGAAGTGACGATGGGGCCAT
TCCGTGGGGCGCGTTACACCAGGCGACTGTCAGTACAGCAATCGAGAGTGGGCTGAT
CAGCCCCTGTGCGTTCTGGCCATCGACGCCTCTTTTCACCGCAAAGCCGGTCAGCA
CACCGCACACCTCGGCTCGTTCTGGAATGGCTGTGCCGCGCGGACC

15 This double stranded DNA loop is bounded on the right by
a T2 control element whose identifier is 2694. This T2
control element has the DNA sequence

Seq. Id. = 113 Position = 1 to 274

20 GCTGAACGCCCTGAATCTCTCCCGGTATGCAGCCTGCTCGGAGAGTACGATTCGTTCG
TTGGCTGCACCGAAGTGACGATGGGGCCATTCCGTGGGGCGCGTTACACCAGGCGAC
TGTCAGTACAGCAATCGAGAGTGGGCTGATCAGCCCCTGTGCGTTCTGGCCATCGA
CGCCTCTTTTCACCGCAAAGCCGGTCAGCACACCGCACACCTCGGCTCGTTCTGGAA
25 TGGCTGTGCCGCGCGGACCGAACGCGGAATCGAGCAATCCTGTTGT

This long T1/T2 double stranded DNA loop modulates the
expression of the following genes

30	DRB0020	DRB0021	DRB0022	DRB0023	DRB0024
	DRB0025	DRB0027	DRB0030	DRB0032	DRB0033
	DRB0034	DRB0035	DRB0037	DRB0038	DRB0039
	DRB0041	DRB0042	DRB0043	DRB0044	DRB0045

DRB0047 DRB0051 DRB0052 DRB0054 DRB0055
DRB0057

This long T1/T2 double stranded DNA loop modulates the
5 expression of the following C1/C2 short loops

A C1/C2 short loop on chromosome 3 (plasmid MP1) whose
identifier is 2693 controls the expression of the genes
of one or more other T1/T2 long loops. This C1/C2 short
10 loop is expressed as a RNA single strand that is 3'UTR to
the gene DRB0057 and has the DNA sequence

Seq. Id. = 114 Position = 1 to 103

15 CTGATGGCCATCCTACAGTACGTTCTCAGCGCGGTCCCGCTGCGCAAGACGCAGCGG
AATTTCCTGACCGTGCTGCTCAGCGTTTTTCTCGCTGTTCTGAC

The expression of genes in this T1/T2 long loop is
controlled by the following C1/C2 short loops.

20

A C1/C2 short loop on chromosome 1 whose identifier is 16
controls the expression of the genes in this T1/T2 long
loop. This C1/C2 short loop is expressed as a RNA single
strand that is 3'UTR to the gene DR0009 and has the DNA
25 sequence

Seq. Id. = 115 Position = 1 to 186

GCTGTGAAATCACCGCTTCCAATGGGTCTGATGGCCATCCTACAGTACGTTCTCAGC
30 GCGGTCCCGCTGCGCAAGACGCAGCGGAATTTCTGACCGTGCTGCTCAGCGTTTTT
CTCGCTGTTCTGACGGCTGAACGCCCTGAATCTCTCCCGGTATGCAGCCTGCTCG
GAGAGTACGATTCGT

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 115 Position = 105 to 186

5

CAGCGTTTTCTCGCTGTTTCCTGGACGGCTGAACGCCCTGAATCTCTCCCGGTATGC
AGCCTGCTCGGAGAGTACGATTCGT

10 The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 115 Position = 132 to 186

15

GCTGAACGCCCTGAATCTCTCCCGGTATGCAGCCTGCTCGGAGAGTACGATTCGT

A C1/C2 short loop on chromosome 3 (plasmid MP1) whose identifier is 2768 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene DRB0133 and has the DNA sequence

20

Seq. Id. = 116 Position = 1 to 186

25 GCTGTGAAATCACCGCTTCCAATGGGTCTGATGGCCATCCTACAGTACGTTCTCAGC
GCGGTCCCGCTGCGCAAGACGCAGCGGAATTCCTGACCGTGCTGCTCAGCGTTTTT
CTCGCTGTTTCCTGGACGGCTGAACGCCCTGAATCTCTCCCGGTATGCAGCCTGCTCG
GAGAGTACGATTCGT

30 The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 116 Position = 105 to 186

CAGCGTTTTTCTCGCTGTTCCCTGGACGGCTGAACGCCCTGAATCTCTCCCGGTATGC
AGCCTGCTCGGAGAGTACGATTCGT

5 The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. = 116 Position = 132 to 186

GCTGAACGCCCTGAATCTCTCCCGGTATGCAGCCTGCTCGGAGAGTACGATTCGT

10

A C1/C2 short loop on chromosome 3 (plasmid MP1) whose
identifier is 2653 controls the expression of the genes
in this T1/T2 long loop. This C1/C2 short loop is
expressed as a RNA single strand that is 3'UTR to the
15 gene DRB0017 and has the DNA sequence

Seq. Id. = 117 Position = 1 to 186

CGGTCCCGCTGCGCAAGACGCAGCGGAATTTCCCTGACCGTGCTGCTCAGCGTTTTTC
20 TCGCTGTTCCCTGGACGGCTGAACGCCCTGAATCTCTCCCGGTATGCAGCCTGCTCGG
AGAGTACGATTCGTTCGTTGGCTGCACCGAAGTGACGATGGGGCCATTCCGTGGGGCG
CGTTACACCAGGCGA

25 The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 117 Position = 47 to 186

CAGCGTTTTTCTCGCTGTTCCCTGGACGGCTGAACGCCCTGAATCTCTCCCGGTATGC
30 AGCCTGCTCGGAGAGTACGATTCGTTCGTTGGCTGCACCGAAGTGACGATGGGGCCAT
TCCGTGGGGCGCGTTACACCAGGCGA

The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 117 Position = 74 to 186

5

GCTGAACGCCCTGAATCTCTCCCGGTATGCAGCCTGCTCGGAGAGTACGATTCGTCG
TTGGCTGCACCGAAGTGACGATGGGGCCATTCCGTGGGGCGCGTTACACCAGGCGA

10

A double stranded DNA loop of length 68.612 kilo-bases on chromosome 3 (plasmid MP1) is bounded on the left by a T1 sequence whose identifier is 2692. This T1 control element has the DNA sequence

15

Seq. Id. = 118 Position = 1 to 103

CTGATGGCCATCCTACAGTACGTTCTCAGCGCGGTCCCGCTGCGCAAGACGCAGCGG
AATTCCTGACCGTGCTGCTCAGCGTTTTTCTCGCTGTTCTGAC

20

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 2749. This T2 control element has the DNA sequence

25

Seq. Id. = 119 Position = 1 to 103

AGCGCGGTCCCGCTGCGCAAGACGCAGCGGAATTCCTGACCGTGCTGCTCAGCGTT
TTTCTCGCTGTTCTGACGGCTGAACGCCCTGAATCTCTCCCGGT

30

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

	DRB0059	DRB0060	DRB0061	DRB0062	DRB0064
	DRB0065	DRB0066	DRB0067	DRB0068	DRB0069
	DRB0070	DRB0072	DRB0073	DRB0074	DRB0076
	DRB0077	DRB0079	DRB0080	DRB0081	DRB0083
5	DRB0085	DRB0086	DRB0087	DRB0088	DRB0089
	DRB0090	DRB0092	DRB0093	DRB0094	DRB0096
	DRB0097	DRB0098	DRB0102	DRB0103	DRB0104
	DRB0105	DRB0106	DRB0107	DRB0111	
	DRB0112				

10

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

15 A C1/C2 short loop on chromosome 3 (plasmid MP1) whose identifier is 2693 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene DRB0057 and has the DNA sequence

20 Seq. Id. = 120 Position = 1 to 103

CTGATGGCCATCCTACAGTACGTTCTCAGCGCGGTCCCGCTGCGCAAGACGCAGCGG
AATTCCTGACCGTGCTGCTCAGCGTTTTTCTCGCTGTTCTCCTGGAC

25 A C1/C2 short loop on chromosome 3 (plasmid MP1) whose identifier is 2695 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene DRB0057 and has the DNA sequence

30

Seq. Id. = 121 Position = 1 to 274

GCTGAACGCCCTGAATCTCTCCCGGTATGCAGCCTGCTCGGAGAGTACGATTCGTCG
 TTGGCTGCACCGAAGTGACGATGGGGCCATTCCGTGGGGCGCGTTACACCAGGCGAC
 TGTCAGTACAGCAATCGAGAGTGGGCTGATCAGCCCACTGTGCGTTCTGGCCATCGA
 CGCCTCTTTTCACCGCAAAGCCGGTCAGCACACCGCACACCTCGGCTCGTTCTGGAA
 5 TGGCTGTGCCGCGCGGACCGAACGCGGAATCGAGCAATCCTGTTGT

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

10 A C1/C2 short loop on chromosome 1 whose identifier is 16
 controls the expression of the genes in this T1/T2 long
 loop. This C1/C2 short loop is expressed as a RNA single
 strand that is 3'UTR to the gene DR0009 and has the DNA
 sequence

15

Seq. Id. = 122 Position = 1 to 186

GCTGTGAAATCACCGCTTCCAATGGGTCTGATGGCCATCCTACAGTACGTTCTCAGC
 GCGGTCCCGCTGCGCAAGACGCAGCGGAATTTTCCTGACCGTGCTGCTCAGCGTTTTT
 20 CTCGCTGTTCTGACGGCTGAACGCCCTGAATCTCTCCCGGTATGCAGCCTGCTCG
 GAGAGTACGATTCGT

The match between the T1 sequence and the C1/C2 sequence is

25

Seq. Id. = 122 Position = 28 to 130

CTGATGGCCATCCTACAGTACGTTCTCAGCGCGGTCCCGCTGCGCAAGACGCAGCGG
 AATTTTCCTGACCGTGCTGCTCAGCGTTTTTCTCGCTGTTCTGAC

30

The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 122 Position = 55 to 157

AGCGCGGTCCCGCTGCGCAAGACGCAGCGGAATTTCTGACCGTGCTGCTCAGCGTT
TTTCTCGCTGTTCTGACGGCTGAACGCCCTGAATCTCTCCCGGT

5

A C1/C2 short loop on chromosome 3 (plasmid MP1) whose identifier is 2768 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene DRB0133 and has the DNA sequence

10

Seq. Id. = 123 Position = 1 to 309

GCTGTGAAATCACCGCTTCCAATGGGTCTGATGGCCATCCTACAGTACGTTCTCAGC
GCGGTCCCGCTGCGCAAGACGCAGCGGAATTTCTGACCGTGCTGCTCAGCGTTTTT
CTCGCTGTTCTGACGGCTGAACGCCCTGAATCTCTCCCGGTATGCAGCCTGCTCG
GAGAGTACGATTCGTCGGACCGAACGCGGAATCGAGCAATCCTGTTGTGCCCTCATT
GATGTCCAGCACCGGCAGGCCTTGACGGTTCGATGTCCGTCAGACCCTGACCGGGTCT
GAGGCTCCAACCTCGTCTGGAACAG

20

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 123 Position = 28 to 130

25

CTGATGGCCATCCTACAGTACGTTCTCAGCGCGGTCCCGCTGCGCAAGACGCAGCGG
AATTTCTGACCGTGCTGCTCAGCGTTTTTCTCGCTGTTCTGAC

The match between the T2 sequence and the C1/C2 sequence is

30

Seq. Id. = 123 Position = 55 to 107

AGCGCGGTCCCGCTGCGCAAGACGCAGCGGAATTCCTGACCGTGCTGCTCAGCGTT
TTTCTCGCTGTTTCCTGGACGGCTGAACGCCCTGAATCTCTCCCGGT

5 A C1/C2 short loop on chromosome 3 (plasmid MP1) whose
identifier is 2693 controls the expression of the genes
in this T1/T2 long loop. This C1/C2 short loop is
expressed as a RNA single strand that is 3'UTR to the
gene DRB0057 and has the DNA sequence

10 Seq. Id. = 124 Position = 1 to 103

CTGATGGCCATCCTACAGTACGTTCTCAGCGCGGTCCCGCTGCGCAAGACGCAGCGG
AATTCCTGACCGTGCTGCTCAGCGTTTTTCTCGCTGTTTCCTGGAC

15 The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 124 Position = 1 to 103

20 CTGATGGCCATCCTACAGTACGTTCTCAGCGCGGTCCCGCTGCGCAAGACGCAGCGG
AATTCCTGACCGTGCTGCTCAGCGTTTTTCTCGCTGTTTCCTGGAC

The match between the T2 sequence and the C1/C2 sequence
is

25

Seq. Id. = 124 Position = 28 to 103

AGCGCGGTCCCGCTGCGCAAGACGCAGCGGAATTCCTGACCGTGCTGCTCAGCGTT
TTTCTCGCTGTTTCCTGGAC

30

A C1/C2 short loop on chromosome 3 (plasmid MP1) whose
identifier is 2653 controls the expression of the genes
in this T1/T2 long loop. This C1/C2 short loop is

expressed as a RNA single strand that is 3'UTR to the gene DRB0017 and has the DNA sequence

Seq. Id. = 125 Position = 1 to 186

5

CGGTCCCGCTGCGCAAGACGCAGCGGAATTCCTGACCGTGCTGCTCAGCGTTTTTC
TCGCTGTTCTGACGGCTGAACGCCCTGAATCTCTCCCGGTATGCAGCCTGCTCGG
AGAGTACGATTCGTCTGTTGGCTGCACCGAAGTGACGATGGGGCCATTCCGTGGGGCG
CGTTACACCAGGCGA

10

The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 125 Position = 1 to 172

15

CGGTCCCGCTGCGCAAGACGCAGCGGAATTCCTGACCGTGCTGCTCAGCGTTTTTC
TCGCTGTTCTGAC

The match between the T2 sequence and the C1/C2 sequence
20 is

Seq. Id. = 125 Position = 1 to 99

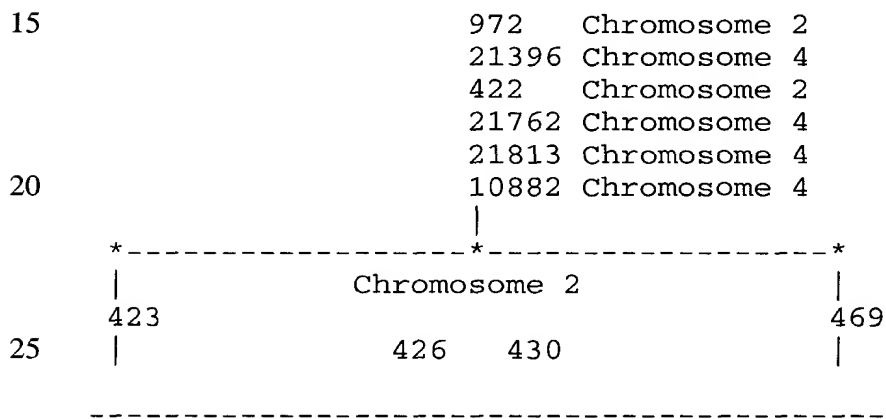
CGGTCCCGCTGCGCAAGACGCAGCGGAATTCCTGACCGTGCTGCTCAGCGTTTTTC
25 TCGCTGTTCTGACGGCTGAACGCCCTGAATCTCTCCCGGT

5. Connectrons occur in plants and higher animals

Connectron relationships exist in plant and higher
5 animals.

Example of a plant connectron - *A. thaliana*

In this example the existence of the T1-T2 (423-469)
10 long loop is controlled by six C1/C2 short loops (972,
21396, 422, 21762, 21813 and 10882). The T1-T2 long loop
controls the expression of six genes on chromosome 2 in
addition to two C1/C2 (426 and 430) short loops.



A double stranded DNA loop of length 42.285 kilo-bases on
30 chromosome 2 is bounded on the left by a T1 sequence
whose identifier is 423. This T1 control element has the
DNA sequence

Seq. Id. = 126 Position = 1 to 67

35

TATCTCTTTAAGGATTAAAAAGTCAAATACTAATTTAATTAATTAAATTTAATTAAA
AAACGAAATA

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 469. This T2 control element has the DNA sequence

5

Seq. Id. = 127 Position = 1 to 67

TACTAATTTAATTAATTAATAATTTAATTAAAAACGAAATACATTATTAATTTTCAAA
AATAATAACC

10

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

At2g02070 At2g02080 At2g02090 At2g02100 At2g02120
15 At2g02130

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

20 A C1/C2 short loop on chromosome 2 whose identifier is 426 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene At2g02060 and has the DNA sequence

25

Seq. Id. = 128 Position = 1 to 55

TTCCAAAATAATAACCAATCAAAATCAACATATAAGATTTGATATCTAAATTTT

30 A C1/C2 short loop on chromosome 2 whose identifier is 430 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is

expressed as a RNA single strand that is 3'UTR to the gene At2g02060 and has the DNA sequence

Seq. Id. = 129 Position = 1 to 55

5

TTGCGGAAAAATAATATCATCATTATAAAAAATAATTAGAGTTTTTTCGCATAT

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

10

A C1/C2 short loop on chromosome 2 whose identifier is 972 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene At2g04240 and has the DNA sequence

15

Seq. Id. = 130 Position = 1 to 118

GTATGCCATTAGAAATAAAATTTTAAAAGTAAATTAATTCATCTCTTTAAAATTAA
20 AAAGTCAAATACTAATTTAATTAATTAATTTAATTAATAACGAAATACATTATTA
ATTT

The match between the T1 sequence and the C1/C2 sequence is

25

Seq. Id. = 130 Position = 53 to 106

ATTAAAAAGTCAAATACTAATTTAATTAATTAATTTAATTAATAACGAAATA

30 The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 130 Position = 167 to 118

TACTAATTTAATTAATTAATAATTTAATTAAAAACGAAATACATTATTAATTT

5 A C1/C2 short loop on chromosome 4 whose identifier is
21396 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene AT4g15300 and has
the DNA sequence

10 Seq. Id. = 131 Position = 1 to 122

TGCCATTAGAAATAAAATTTTAAAGAGTAAATTAATTTATCTCTTTAAGGATTAAAA
AGTCAAATACTAATTTAATTAATTAATAATTTAATTAAAAACGAAATACATTATTAAT
TTCCAAAA

15

The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 131 Position = 38 to 104

20

TATCTCTTTAAGGATTAAAAAGTCAAATACTAATTTAATTAATTAAATTTAATTAAA
AAACGAAATA

25 The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. = 131 Position = 65 to 116

TACTAATTTAATTAATTAATAATTTAATTAAAAACGAAATACATTATTAATTT

30

A C1/C2 short loop on chromosome 2 whose identifier is
422 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA

single strand that is 3'UTR to the gene At2g02060 and has the DNA sequence

Seq. Id. = 132 Position = 1 to 137

5

TAACCTTAATTTTTGTAAGTAATTATATAGGTATGCCATTAGAAATAAAATTTTAAA
GAGTAAATTAATTTATCTCTTTAAGGATTAAAAAGTCAAATACTAATTTAATTAATT
AAATTTAATTAAAAACGAAATA

10 The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 132 Position = 71 to 137

15 TATCTCTTTAAGGATTAAAAAGTCAAATACTAATTTAATTAATTAAATTTAATTAAA
AAACGAAATA

The match between the T2 sequence and the C1/C2 sequence is

20

Seq. Id. = 132 Position = 98 to 137

TACTAATTTAATTAATTAAATTTAATTAAAAACGAAATA

25 A C1/C2 short loop on chromosome 4 whose identifier is 21762 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene AT4g17510 and has the DNA sequence

30

Seq. Id. = 133 Position = 1 to 65

TTTAAGGATTAAAAAGTCAAATACTAATTTAATTAATTAAATTTAATTAAAAACGA
AATACATT

5 The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 133 Position = 1 to 61

10 TTTAAGGATTAAAAAGTCAAATACTAATTTAATTAATTAAATTTAATTAAAAACGA
AATA

The match between the T2 sequence and the C1/C2 sequence
is

15 Seq. Id. = 133 Position = 22 to 65

TACTAATTTAATTAATTAAATTTAATTAAAAACGAAATACATT

20 A C1/C2 short loop on chromosome 4 whose identifier is
21813 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene AT4g17680 and has
the DNA sequence

25 Seq. Id. = 134 Position = 1 to 65

TTTAAGGATTAAAAAGTCAAATACTAATTTAATTAATTAAATTTAATTAAAAACGA
AATACATT

30 The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 134 Position = 1 to 61

TTTAAGGATTAAAAAGTCAAATACTAATTTAATTAATTAAATTTAATTAAAAACGA
AATA

- 5 The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. = 134 Position = 22 to 65

- 10 TACTAATTTAATTAATTAAATTTAATTAAAAACGAAATACATT

- A C1/C2 short loop on chromosome 2 whose identifier is
10882 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
15 single strand that is 3'UTR to the gene At2g26540 and has
the DNA sequence

Seq. Id. = 135 Position = 1 to 56

- 20 TATCTCTTTAAGGATTAAAAAGTCAAATACTAATTTAATTAATTAAATTTAATTAA

The match between the T1 sequence and the C1/C2 sequence
is

- 25 Seq. Id. = 135 Position = 1 to 56

TATCTCTTTAAGGATTAAAAAGTCAAATACTAATTTAATTAATTAAATTTAATTAA

- The match between the T2 sequence and the C1/C2 sequence
30 is

Seq. Id. = 135 Position = 28 to 56

TACTAATTTAATTAATTAAATTTAATTAA

5 Example of a animal connectron - D. megalomaster

A double stranded DNA loop of length 88.159 kilo-bases on
chromosome 4 is bounded on the left by a T1 sequence
whose identifier is 3340. This T1 control element has
10 the DNA sequence

Seq. Id. = 136 Position = 1 to 132

ACCTAAAAGAAGTACCGTTTTTTACTCCTAATTACCAATTCTAACCATCCATATCAC
15 TTTTTGACGGACTCCGTGAAAATAATTTTTGGCCAAATTTTCGCATTTTTTGTAAGG
GGTAACATCATAAAAATT

This double stranded DNA loop is bounded on the right by
a T2 control element whose identifier is 3372. This T2
20 control element has the DNA sequence

Seq. Id. = 137 Position = 1 to 136

AAAAAAGTACCGCGTTTTACTCCTAATTACCAATTCTAACCATCCATATCACTTTTTT
25 GACGGACTCCGTGAAAATAATTTTTGGCCAAATTTTCGCATTTTTTGTAAGGGGTAA
CATCATCAAAATTTGCGAAAAA

This long T1/T2 double stranded DNA loop modulates the
expression of the following genes

30

[Some of the following gene names have not been
determined.]

-	-	-	-	-
-	CG11207	-	CG2186	CG2157
-	Ork1	-	-	-
-				

5

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

10 A C1/C2 short loop on chromosome 4 whose identifier is 3362 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene XXX and has the DNA sequence

15 Seq. Id. = 138 Position = 1 to 134

AAAAAAGTACCGCGTTTTACTCCTAATTACCAATTCTAACCATCCATATCACTTTTT
GACGGACTCCGTTAAAATAATTTTTGACCAAATTTTCGCATTTTTTGTAATCAAAAT
TTGCAAAAAATTGAAAAAC

20

A C1/C2 short loop on chromosome 4 whose identifier is 3364 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the
25 gene XXX and has the DNA sequence

Seq. Id. = 139 Position = 1 to 83

30 CAAATTTGAATGCAAATCGATTGGGAATCAAAAAACAACTCAACGAGGTATGACA
TTCCATATTTGGGCCATTATTTCCAA

A C1/C2 short loop on chromosome 4 whose identifier is 3366 controls the expression of the genes of one or more

other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene XXX and has the DNA sequence

5 Seq. Id. = 140 Position = 1 to 62

TTTTTTCACAAAAATTAGGAAAATGATTTTGGGTAAAAAATGAATATTTAAGTTGG
GTTTT

10 A C1/C2 short loop on chromosome 4 whose identifier is 3369 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene XXX and has the DNA sequence

15

Seq. Id. = 141 Position = 1 to 87

AAATCGATTGGGAATCAAAAAACAAACCTCAACGAGGTATGACATTCCATATCTGGG
CCATTATTTCCAATCTTTTGATCAAAATAC

20

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 4 whose identifier is
25 3373 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene XXX and has the DNA sequence

30 Seq. Id. = 142 Position = 1 to 136

AAAAAAGTACCGCGTTTTACTCCTAATTACCAATTCTAACCATCCATATCACTTTTT
GACGGACTCCGTGAAAATAATTTTTGGCCAAATTTTCGCATTTTTTGTAAGGGGTAA
CATCATCAAATTTGCGAAAAA

5 The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 142 Position = 15 to 120

10 TTTTACTCCTAATTACCAATTCTAACCATCCATATCACTTTTTGACGGACTCCGTGA
AAATAATTTTTGGCCAAATTTTCGCATTTTTTGTAAGGGGTAAACATCAT

The match between the T2 sequence and the C1/C2 sequence
is

15

Seq. Id. = 142 Position = 1 to 136

AAAAAAGTACCGCGTTTTACTCCTAATTACCAATTCTAACCATCCATATCACTTTTT
GACGGACTCCGTGAAAATAATTTTTGGCCAAATTTTCGCATTTTTTGTAAGGGGTAA
20 CATCATCAAATTTGCGAAAAA

Example of an animal connectron - H. sapiens

25

All of the human genome that has been fully sequenced by
both the NIH-lead global sequencing project and the
Celera Genomics, Inc. project. The gene descriptors for
this chromosome do not yet exist. Without the positions
30 and directions of the genes, it is not possible to select
from among the possible connectrons to determine the real
connectrons.

Human chromosome 22 has been processed and there 31,000 possible connectrons.

5 The gene descriptors for all the chromosomes of the human genome should become available within the year.

6. Permanent connectrons exist in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes.

5

C1/C2 short loops are normally expressed as the 3'UTR of some gene. A class of connectron relationships exist that permit one C1/C2 short loop to control the existence of one or more T1-T2 long loops without being subject to any expression controls other than those of the gene to which the C1/C2 is 3'UTR. These connectron relationships are described as "permanent". Permanent connectrons exist in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes.

10

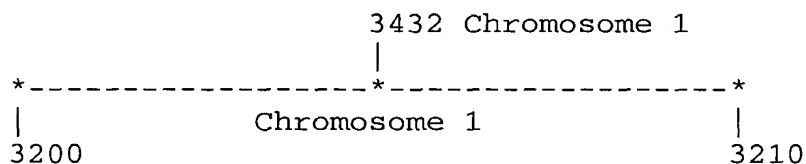
15

Example of a prokaryote permanent connectron - E. coli

In this example the existence of the T1-T2 (3200-3210) long loop is controlled by a C1/C2 short loop (3432). The expression of this C1/C2 short loop is controlled only by the gene btuB.

20

25



30

A double stranded DNA loop of length 93.339 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 3200. This T1 control element has the DNA sequence

35

Seq. Id. = 143 Position = 1 to 378

AAGCGGCACTGCTCTTTAACAATTTATCAGACAATCTGTGTGGGCACTCGAAGATAC
GGATTCTTAACGTCGCAAGACGAAAAATGAATACCAAGTCTCAAGAGTGAACACGTA
5 ATTCATTACGAAGTTTAATTCTTTGAGCATCAAACCTTTTAAATTGAAGAGTTTGATC
ATGGCTCAGATTGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAACGGTAACAG
GAAACAGCTTGCTGTTTCGCTGACGAGTGGCGGACGGGTGAGTAATGTCTGGGAAAC
TGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATACCGCATAACGTCGCA
AGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATC

10

This double stranded DNA loop is bounded on the right by
a T2 control element whose identifier is 3310. This T2
control element has the DNA sequence

15 Seq. Id. = 144 Position = 1 to 378

CAGACAATCTGTGTGGGCACTCGAAGATACGGATTCTTAACGTCGCAAGACGAAAAA
TGAATACCAAGTCTCAAGAGTGAACACGTAATTCATTACGAAGTTTAATTCTTTGAG
CGTCAAACCTTTTAAATTGAAGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAG
20 GCCTAACACATGCAAGTCGAACGGTAACAGGAAGAAGCTTGCTTCTTTGCTGACGAG
TGGCGGACGGGTGAGTAATGTCTGGGAACTGCCTGATGGAGGGGGATAACTACTGG
AAACGGTAGCTAATACCGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCGGGCCT
CTTGCCATCGGATGTGCCCAGATGGGATTAGCTAGT

25 This long T1/T2 double stranded DNA loop modulates the
expression of the following genes

rrsC	gltU	rrlC	rrfC	aspT
trpT	yifA	yifE	yifB	ilvL
30 ilvG_1	ilvM	ilvE	ilvD	ilvA
ilvY	ilvC	ppiC	b3776	rep
gppA	rhlB	trxA	rhoL	rho
rfe	wzzE	wecB	rffH	wecD

	wecE	wzxE	yifM_2	wecG	yifK
	argX	hisR	leuT	proM	aslB
	aslA	hemY	hemX	hemD	cyaA
	cyaY	b3808	dapF	uvrD	b3814
5	corA	yigF	yigG	rarD	yigI
	pldA	recQ	yigJ	yigK	pldB
	yigL	yigM	metR	metE	ysgA
	udp	yigN	ubiE	yigP	b3836
	yigU	yigW_1	rfaH	yigC	ubiB
10	fadA	fadB	pepQ		trkH
	hemG				

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

15

A C1/C2 short loop on chromosome 1 whose identifier is 3432 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene btuB and has the

20 DNA sequence

Seq. Id. = 145 Position = 1 to 520

AAGCGGCACTGCTCTTTAACAATTTATCAGACAATCTGTGTGGGCACTCGAAGATAC
 25 GGATTCTTAACGTCGCAAGACGAAAAATGAATACCAAGTCTCAAGAGTGAACACGTA
 ATTCATTACGAAGTTTAATTCTTTGAGCCAGACAATCTGTGTGGGCACTCGAAGATA
 CGGATTCTTAACGTCGCAAGACGAAAAATGAATACCAAGTCTCAAGAGTGAACACGT
 AATTCATTACGAAGTTTAATTCTTTGAGCGTCAAACCTTTTAAATTGAAGAGTTTGAT
 CATGGCTCAGATTGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAACGGTAACA
 30 GGAAGAAGCTTGCTTCTTTGCTGACGAGTGGCGGACGGGTGAGTAATGTCTGGGAAA
 CTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATACCGCATAACGTTCG
 AAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCGGATGTGCCAGATGGGATT
 AGCTAGT

The match between the T1 sequence and the C1/C2 sequence is

5 Seq. Id. = 145 Position = 1 to 142

AAGCGGCACTGCTCTTTAACAATTTATCAGACAATCTGTGTGGGCACTCGAAGATAC
GGATTCTTAACGTCGCAAGACGAAAAATGAATACCAAGTCTCAAGAGTGAACACGTA
ATTCATTACGAAGTTTAATTCTTTGAGC

10

The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 145 Position = 143 to 520

15

CAGACAATCTGTGTGGGCACTCGAAGATACGGATTCTTAACGTCGCAAGACGAAAAA
TGAATACCAAGTCTCAAGAGTGAACACGTAATTCATTACGAAGTTTAATTCTTTGAG
CGTCAAACCTTTTAAATTGAAGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAG
GCCTAACACATGCAAGTCGAACGGTAACAGGAAGAAGCTTGCTTCTTTGCTGACGAG
20 TGGCGGACGGGTGAGTAATGTCTGGGAACTGCCTGATGGAGGGGGATAACTACTGG
AAACGGTAGCTAATACCGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCGGGCCT
CTTGCCATCGGATGTGCCCAGATGGGATTAGCTAGT

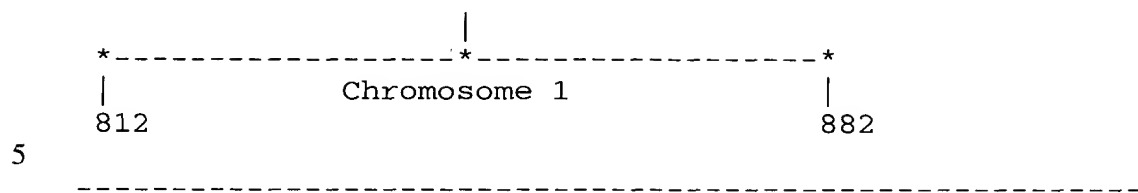
25

Example of an archea permanent connectron - *H. pylori*

In this example the existence of the T1-T2 (812-882) long loop is controlled by a C1/C2 short loop (1241).

30 The expression of this C1/C2 short loop is controlled only by the gene HP1535.

1241 Chromosome 1



A double stranded DNA loop of length 96.385 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 812. This T1 control element has the DNA sequence

Seq. Id. = 146 Position = 1 to 43

15 TTTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCA

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 882. This T2 control element has the DNA sequence

20

Seq. Id. = 147 Position = 1 to 43

TAGCGGAACTAAAGCATTCATCCCAAACACTAAAGATATTTGG

25 This long T1/T2 double stranded DNA loop modulates the expression of the following genes

HP0999	HP1000	HP1001	HP1002	HP1003
HP1005	HP1006	HP1008	HP1009	HPtRNA-Pro
30 HP1010	HP1011	HP1013	HP1015	HP1017
HP1018	HP1020	HP1021	HP1022	HP1023
HP1024	HP1025	HP1027	HP1028	HP1030
HP1031	HP1033	HP1034	HP1038	HP1039
HP1040	HP1041	HP1042	HP1043	HP1044
35 HP1045	HP1046	HP1051	HP1052	HP1055

HP1056	HP1058	HP1060	HP1065	HPtRNA-Ser
HP1066	HP1067	HP1069	HP1070	HP1074
HP1075	HP1076	HP1077	HP1078	HP1079
HP1080	HP1081	HP1083	HP1084	HP1085
5 HP1088	HP1091	HP1092	HP1093	HP1094
HP1095	HP1096			

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

10

A C1/C2 short loop on chromosome 1 whose identifier is 1241 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene HP1535 and has
15 the DNA sequence

Seq. Id. = 148 Position = 1 to 56

TTTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCATTTCATCCCAAACA

20

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 148 Position = 1 to 43

25

TTTTACTCATAGGGTTTTTATAGTTCCTAGCGGAACTAAAGCA

The match between the T2 sequence and the C1/C2 sequence is

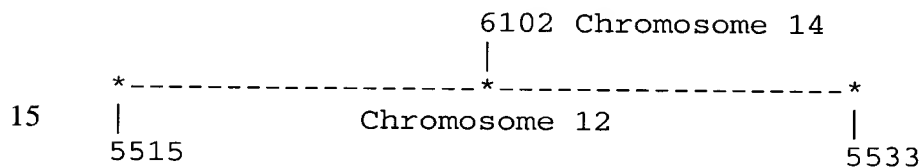
30

Seq. Id. = 148 Position = 28 to 56

TAGCGGAACTAAAGCATTTCATCCCAAACA

Example of a single-celled permanent connectron - S.
5 cervesiae

In this example the existence of the T1-T2 (5515-5533)
long loop is controlled by a C1/C2 short loop (6102).
The expression of this C1/C2 short loop is controlled
10 only by the gene YNL339C.



20 A double stranded DNA loop of length 6.466 kilo-bases on
chromosome 12 is bounded on the left by a T1 sequence
whose identifier is 5515. This T1 control element has
the DNA sequence

25 Seq. Id. = 149 Position = 1 to 225

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
AGAAATTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
30 AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTG

This double stranded DNA loop is bounded on the right by
a T2 control element whose identifier is 5533. This T2
control element has the DNA sequence

35

Seq. Id. = 150 Position = 1 to 225

ATTATGTATTGTGTAGTATAGTATATTGTAAGAAATTTTTTTTCTAGGGAATATGC
GTTTTGATGTAGTAGTATTTCACTGTTTTGATTTAGTGTTTGTGTCACGGCAGTAGC
5 GAGAGACAAGTGGGAAAGAGTAGGATAAAAAGACAATCTATAAAAAGTAAACATAAA
ATAAAGGTAGTAAGTAGCTTTTGGTTGAACATCCGGGTAAGAGACAACAGGGCT

This long T1/T2 double stranded DNA loop modulates the
expression of the following genes

10

YLR467W

The expression of genes in this T1/T2 long loop is
controlled by the following C1/C2 short loops.

15

A C1/C2 short loop on chromosome 14 whose identifier is
6102 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene YNL339C and has
20 the DNA sequence

Seq. Id. = 151 Position = 1 to 252

AGGAAATTGTTGTTACGAAAGTCAGTGATTATGTATTGTGTAGTATAGTATATTGTA
25 AGAAATTTTTTTTTTCTAGGGAATATGCGTTTTGATGTAGTAGTATTTCACTGTTTTG
ATTTAGTGTTTGTGTCACGGCAGTAGCGAGAGACAAGTGGGAAAGAGTAGGATAAAA
AGACAATCTATAAAAAGTAAACATAAAATAAAGGTAGTAAGTAGCTTTTGGTTGAAC
ATCCGGGTAAGAGACAACAGGGCT

30 The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 151 Position = 1 to 225

A double stranded DNA loop of length 30.606 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 569. This T1 control element has the DNA sequence

5

Seq. Id. = 152 Position = 1 to 239

AAATCGAGCCCGTAAATCGACACAAGCGCTACAGTAGTC

10 This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 596. This T2 control element has the DNA sequence

Seq. Id. = 153 Position = 1 to 42

15

AGTGCTACAGTAGTCATTTAAAGAATTACTGTAGTTTTTCGCT

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

20

A C1/C2 short loop on chromosome 5 whose identifier is 24442 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene F20D6.4 and has

25 the DNA sequence

Seq. Id. = 154 Position = 1 to 58

GAGCCCGTAAATCGACACAAGCGCTACAGTAGTCATTTAAAGAATTACTGTAGTTTT

30 C

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 154 Position = 1 to 34

GAGCCCGTAAATCGACACAAGCGCTACAGTAGTC

5

The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. = 154 Position = 23 to 58

10

GCTACAGTAGTCATTTAAAGAATTACTGTAGTTTTC

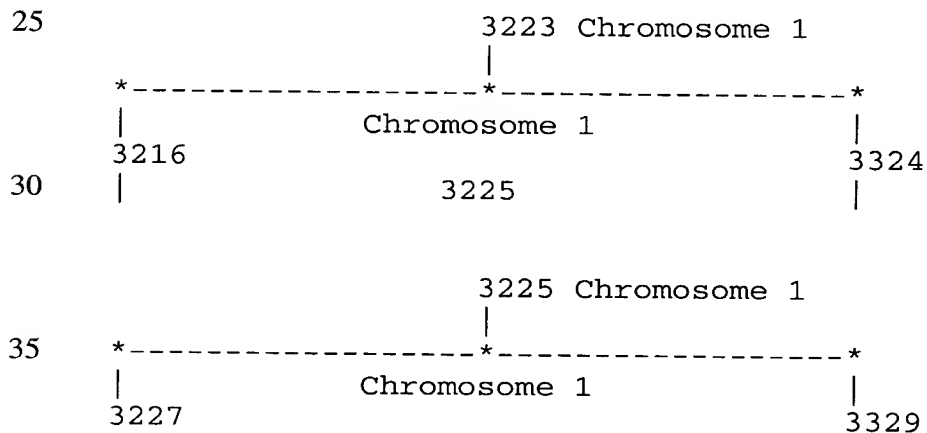
7. Transient connectrons exist in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes.

5

A class of connectron relationships exist that permit one C1/C2 short loop to control the existence of one or more T1-T2 long loops such that this C1/C2 short loop is itself subject to expression control by another T1-T2 long loop which surrounds it. These connectron relationships are described as "transient". Transient connectrons exist in prokaryotes, archaea, single-celled eukaryotes and multi-celled eukaryotes.

15 Example of a prokaryote transient connectron - E. coli

In this example the existence of the T1-T2 (3227-3329) long loop is controlled by the C1/C2 (3225) short loop. The expression of this C1/C2 short loop is controlled by the existence of the T1-T2 (3216-3324) long loop. The existence of this T1-T2 long loop is itself determined by the expression of the C1/C2 (3223) short loop. The C1/C2 (3225) short loop is the transient connectron.



A double stranded DNA loop of length 93.464 kilo-bases on
5 chromosome 1 is bounded on the left by a T1 sequence
whose identifier is 3216. This T1 control element has
the DNA sequence

Seq. Id. = 155 Position = 1 to 337

10

AGCGCAAGCGAAGCTCTTGATCGAAGCCCCGGTAAACGGCGGCCGTAAC TATAACGG
TCCTAAGGTAGCGAAATTCCTTGTCGGGTAAAGTTCCGACCTGCACGAATGGCGTAAT
GATGGCCAGGCTGTCTCCACCCGAGACTCAGTGAAATTGAACTCGCTGTGAAGATGC
AGTGTACCCGCGGCAAGACGGAAAGACCCCGTGAACCTTTACTATAGCTTGACACTG
15 AACATTGAGCCTTGATGTGTAGGATAGGTGGGAGGCTTTGAAGTGTGGACGCCAGTC
TGCATGGAGCCGACCTTGAAATACCACCCTTTAATGTTTGATGTTCTAACGT

This double stranded DNA loop is bounded on the right by
a T2 control element whose identifier is 3324. This T2
20 control element has the DNA sequence

Seq. Id. = 156 Position = 1 to 337

CCCCGTAAACGGCGGCCGTAAC TATAACGGTCCTAAGGTAGCGAAATTCCTTGTCGG
25 GTAAGTTCCGACCTGCACGAATGGCGTAATGATGGCCAGGCTGTCTCCACCCGAGAC
TCAGTGAAATTGAACTCGCTGTGAAGATGCAGTGTACCCGCGGCAAGACGGAAAGAC
CCCGTGAACCTTTACTATAGCTTGACACTGAACATTGAGCCTTGATGTGTAGGATAG
GTGGGAGGCTTTGAAGTGTGGACGCCAGTCTGCATGGAGCCGACCTTGAAATACCAC
CCTTTAATGTTTGATGTTCTAACGTTGACCCGTAATCCGGGTTGCGGACAGT

30

This long T1/T2 double stranded DNA loop modulates the
expression of the following genes

	rrfC	aspT	trpT	yifA	yifE
	yifB	ilvL	ilvG_1	ilvM	ilvE
	ilvD	ilvA	ilvY	ilvC	ppiC
	b3776	rep	gppA	rhlB	trxA
5	rhoL	rho	rfe	wzzE	wecB
	rffH	wecD	wecE	wzxE	yifM_2
	wecG	yifK	argX	hisR	leuT
	proM	aslB	aslA	hemY	hemX
	hemD	cyaA	cyaY	b3808	dapF
10	uvrD	b3814	corA	yigF	yigG
	rarD	yigI	pldA	recQ	yigJ
	yigK	pldb	yigL	yigM	metR
	metE	ysgA	udp	yigN	ubiE
	yigP	b3836	yigU	yigW_1	rfaH
15	yigC	ubiB	fadA	fadB	pepQ
	trkH	hemG	rrsA	ileT	rrlA

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

20

A C1/C2 short loop on chromosome 1 whose identifier is 3225 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the

25

gene rrlC and has the DNA sequence

Seq. Id. = 157 Position = 1 to 137

30

AAACAGAATTTGCCTGGCGGCCGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAAC
TCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTA
GGGAAC TGCCAGGCATCAAATTA

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

5 A C1/C2 short loop on chromosome 1 whose identifier is 3323 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene rrlA and has the DNA sequence

10 Seq. Id. = 158 Position = 1 to 362

GCGAAGCTCTTGATCGAAGCCCCGGTAAACGGCGGCCGTAACCTATAACGGTCCTAAG
GTAGCGAAATTCCTTGTCGGGTAAGTTCCGACCTGCACGAATGGCGTAATGATGGCC
AGGCTGTCTCCACCCGAGACTCAGTGAAATTGAACTCGCTGTGAAGATGCAGTGTAC
15 CCGCGGCAAGACGGAAAGACCCCGTGAACCTTTACTATAGCTTGACACTGAACATTG
AGCCTTGATGTGTAGGATAGGTGGGAGGCTTTGAAGTGTGGACGCCAGTCTGCATGG
AGCCGACCTTGAAATACCACCCTTTAATGTTTGATGTTCTAACGTAACGTTGACCCG
TAATCCGGGTTGCGGACAGT

20 The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 158 Position = 1 to 330

25 GCGAAGCTCTTGATCGAAGCCCCGGTAAACGGCGGCCGTAACCTATAACGGTCCTAAG
GTAGCGAAATTCCTTGTCGGGTAAGTTCCGACCTGCACGAATGGCGTAATGATGGCC
AGGCTGTCTCCACCCGAGACTCAGTGAAATTGAACTCGCTGTGAAGATGCAGTGTAC
CCGCGGCAAGACGGAAAGACCCCGTGAACCTTTACTATAGCTTGACACTGAACATTG
AGCCTTGATGTGTAGGATAGGTGGGAGGCTTTGAAGTGTGGACGCCAGTCTGCATGG
30 AGCCGACCTTGAAATACCACCCTTTAATGTTTGATGTTCTAACGT

The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 158 Position = 21 to 362

5 CCCGGTAAACGGCGGCCGTAAC TATAACGGTCCTAAGGTAGCGAAATTCCTTGTCGG
GTAAGTTCCGACCTGCACGAATGGCGTAATGATGGCCAGGCTGTCTCCACCCGAGAC
TCAGTGAAATTGAACTCGCTGTGAAGATGCAGTGTACCCGCGGCAAGACGGAAAGAC
CCCGTGAACCTTTACTATAGCTTGACACTGAACATTGAGCCTTGATGTGTAGGATAG
GTGGGAGGCTTTGAAGTGTGGACGCCAGTCTGCATGGAGCCGACCTTGAAATACCAC
CCTTTAATGTTTGATGTTCTAACGTTGACCCGTAATCCGGGTTGCGGACAGT

10

15 A double stranded DNA loop of length 93.749 kilo-bases on
chromosome 1 is bounded on the left by a T1 sequence
whose identifier is 3227. This T1 control element has
the DNA sequence

Seq. Id. = 159 Position = 1 to 52

20 AGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGG

This double stranded DNA loop is bounded on the right by
a T2 control element whose identifier is 3329. This T2
control element has the DNA sequence

25

Seq. Id. = 160 Position = 1 to 52

CATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCG

30 This long T1/T2 double stranded DNA loop modulates the
expression of the following genes

	aspT	trpT	yifA	yifE	yifB
	ilvL	ilvG_1	ilvM	ilvE	ilvD
	ilvA	ilvY	ilvC	ppiC	b3776
	rep	gppA	rh1B	trxA	rhoL
5	rho	rfe	wzzE	wecB	rffH
	wecD	wecE	wzxE	yifM_2	wecG
	yifK	argX	hisR	leuT	proM
	aslB	aslA	hemY	hemX	hemD
	cyaA	cyaY	b3808	dapF	uvrD
10	b3814	corA	yigF	yigG	rard
	yigI	pldA	recQ	yigJ	yigK
	pldB	yigL	yigM	metR	metE
	ysgA	udp	yigN	ubiE	yigP
	b3836	yigU	yigW_1	rfaH	yigC
15	ubiB	fadA	fadB	pepQ	trkH
	hemG	rrsA	ileT	rrlA	rrfA

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

20

A C1/C2 short loop on chromosome 1 whose identifier is 3225 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene rrlC and has the

25 DNA sequence

Seq. Id. = 161 Position = 1 to 137

30 AAACAGAATTTGCCTGGCGGCCGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAAC
TCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTA
GGGAAGTGCCAGGCATCAAATTA

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 161 Position = 76 to 127

5

AGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAAGTGCCAGG

The match between the T2 sequence and the C1/C2 sequence is

10

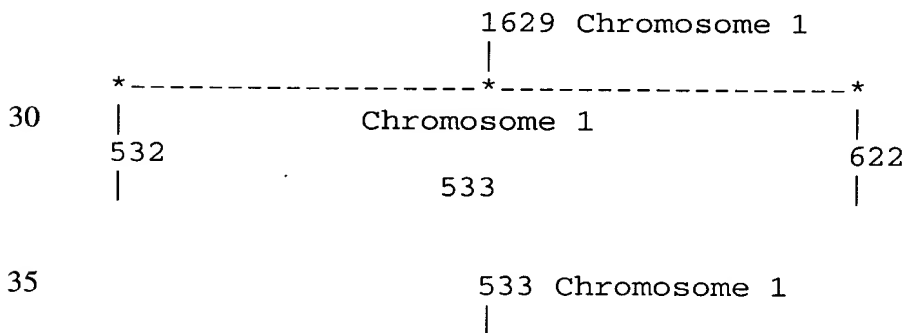
Seq. Id. = 161 Position = 103 to 135

CATGCGAGAGTAGGGAAGTGCCAGGCATCAAAT

15

Example of an archea transient connectron - *M. jannaschii*

In this example the existence of the T1-T2 (1139-1159) long loop is controlled by the C1/C2 (533) short loop. The expression of this C1/C2 short loop is controlled by the existence of the T1-T2 (532-622) long loop. The existence of this T1-T2 long loop is itself determined by the expression of the C1/C2 (1629) short loop. The C1/C2 (533) short loop is the transient connectron.



	MJ0540	MJ0541	MJ0542	MJ0543	MJ0544
	MJ0545	MJ0547	MJ0548	MJ0549	MJ0550
	MJ0552	MJ0553	MJ0554	MJ0555	MJ0556
	MJ0557	MJ0558	MJ0559	MJ0560	MJ0561
5	MJ0562	MJ0563	MJ0564	MJ0565	MJ0566
	MJ0568	MJ0569	MJ0570		

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

10

A C1/C2 short loop on chromosome 1 whose identifier is 533 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene MJ0485 and has the DNA sequence

15

Seq. Id. = 164 Position = 1 to 64

ATTTTTATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTA
TTGAATT

20

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

25 A C1/C2 short loop on chromosome 1 whose identifier is 1629 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene MJ1597 and has the DNA sequence

30

Seq. Id. = 165 Position = 1 to 139

ATATGTTTGAAATTTGAAAATAAGAGTATTTAGAAGTTATTAATTAGTTCAAAGGAT
TTTTATTTAATTTCTAAGGGTTTGCTGGTTTGATTATTTAGAATATTTGAGTTTATT
GAATTATTCAGATTTTAAAAATTA

5 The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 165 Position = 1 to 33

10 ATATGTTTGAAATTTGAAAATAAGAGTATTTAG

The match between the T2 sequence and the C1/C2 sequence
is

15 Seq. Id. = 165 Position = 33 to 60

ATTTAGAAGTTATTAATTAGTTCAAAGGATTTT

20

A double stranded DNA loop of length 14.509 kilo-bases on
chromosome 1 is bounded on the left by a T1 sequence
whose identifier is 1139. This T1 control element has
the DNA sequence

25

Seq. Id. = 166 Position = 1 to 78

ATTTATTAATTAGTTCAAAGGATTTTATTTAATTTCTAAGGGTTAGCTGGTTTGAT
TGTTTAAAATATTTGAGTTTA

30

This double stranded DNA loop is bounded on the right by
a T2 control element whose identifier is 1159. This T2
control element has the DNA sequence

Seq. Id. = 167 Position = 1 to 78

ATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAAT
5 TATTCAGATTTTAAAAATTA

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

10	MJ1096	MJ1097	tRNA-Arg-3	MJ1098	MJ1099
	MJ1100	MJ1101	MJ1102	MJ1103	MJ1104
	MJ1105	MJ1106	MJ1107	MJ1108	

The expression of genes in this T1/T2 long loop is
15 controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 1 whose identifier is 533 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA
20 single strand that is 3'UTR to the gene MJ0485 and has the DNA sequence

Seq. Id. = 168 Position = 1 to 64

25 ATTTTATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAATT

The match between the T1 sequence and the C1/C2 sequence is

30

Seq. Id. = 168 Position = 1 to 37

ATTTTATTTAATTTCTAAGGGTTAGCTGGTTTGATT

The match between the T2 sequence and the C1/C2 sequence is

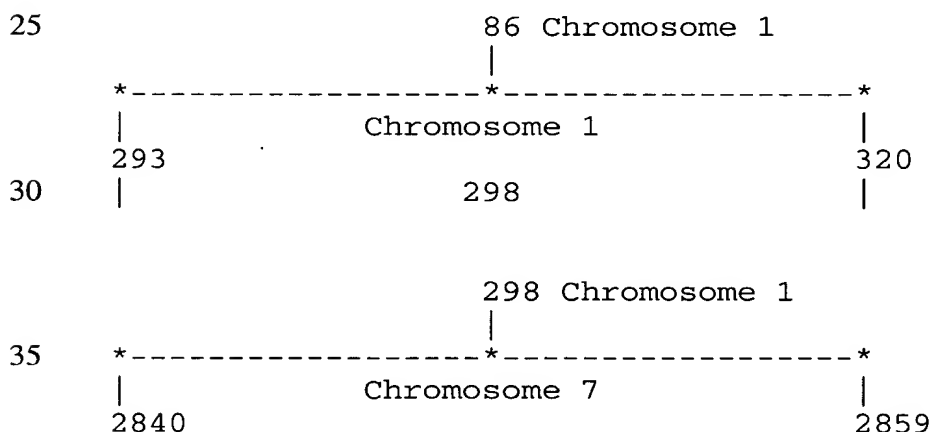
5 Seq. Id. = 168 Position = 7 to 64

ATTTAATTTCTAAGGGTTAGCTGGTTTGATTATTTAGAATATTTGAGTTTATTGAAT
T

10 -----

Example of a single-celled transient connectron - *S. cerevisiae*

15 In this example the existence of the T1-T2 (2840-2859) long loop is controlled by the C1/C2 (298) short loop. The expression of this C1/C2 short loop is controlled by the existence of the T1-T2 (293-320) long loop. The existence of this T1-T2 long loop is itself determined by
20 the expression of the C1/C2 (86) short loop. The C1/C2 (298) short loop is the transient connectron.



A double stranded DNA loop of length 38.470 kilo-bases on
 5 chromosome 2 is bounded on the left by a T1 sequence
 whose identifier is 293. This T1 control element has the
 DNA sequence

Seq. Id. = 169 Position = 1 to 258

10

GAATTGTTGGAATAAAAATCCACTATCGTCTATCAACTAATAGTTATATTATCAATA
 TATTATCATATACGGTGTTAAGATGATGACATAAGTTATGAGAAGCTGTCATCGAAG
 TTAGAGGAAGCTGAAGTGCAAGGATTGATAATGTAATAGGATAATGAAACATATAAA
 ACGGAATGAGGAATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTTGAGG
 15 ATTCCTATATCCTTGAGGAGAACTTCTAGT

This double stranded DNA loop is bounded on the right by
 a T2 control element whose identifier is 320. This T2
 control element has the DNA sequence

20

Seq. Id. = 170 Position = 1 to 70

AATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCTATATCCTCGAGGAG
 AACTTCTAGTATATTCTGTA

25

This long T1/T2 double stranded DNA loop modulates the
 expression of the following genes

YBL005W-B	TS(AGA)B	YBL004W	YBL003C	YBL002W
30 YBL001C	YBR001C	YBR002C	YBR003W	YBR004C
YBR005W	YBR006W	YBR007C	YBR008C	YBR009C
YBR010W	YBR011C	YBR012C		

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

5 A C1/C2 short loop on chromosome 2 whose identifier is 298 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YBL005W-B and has the DNA sequence

10 Seq. Id. = 171 Position = 1 to 342

ATCTATTACATTATGGGTGGTATGTTGGAATAAAAATCCACTATCGTCTATCAACTA
ATAGTTATATTATCAATATATTATCATATACGGTGTTAAGATGATGACATAAGTTAT
GAGAAGCTGTCATCGAAGTTAGAGGAAGCTGAAGTGCAAGGATTGATAATGTAATAG
15 GATAATGAAACATATAAAACGGAATGAGGAATAATCGTAATATTAGTATGTAGAAAT
ATAGATTCCATTTTGAGGATTCCTATATCCTTGAGGAGAACTTCTAGTATATTCTGT
ATACCTAATATTATAGCCTTTATCAACAATGGAATCCCAACAATTATCTCAACATTC

20 The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 1 whose identifier is 86 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single
25 strand that is 3'UTR to the gene YAR009C and has the DNA sequence

Seq. Id. = 172 Position = 1 to 362

30 ATCTATTACATTATGGGTGGTATGTTGGAATAGAAATCAACTATCATCTACTAACTA
GTATTTACATTACTAGTATATTATCATATACGGTGTTAGAAGATGACGCAAATGATG
AGAAATAGTCATCTAAATTAGTGGAAGCTGAAACGCAAGGATTGATAATGTAATAGG
ATCAATGAATATAAACATATAAAACGGAATGAGGAATAATCGTAATATTAGTATGTA

GAAATATAGATTCCATTTTGAGGATTCCTATATCCTCGAGGAGAACTTCTAGTATAT
TCTGTATACCTAATATTATAGCCTTTATCAACAATGGAATCCCAACAATTATCTCAA
CATTCACCCATTTCTCAGAA

5 The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 172 Position = 184 to 264

10 AAACATATAAAACGGAATGAGGAATAATCGTAATATTAGTATGTAGAAATATAGATT
CCATTTTGAGGATTCCTATATCCT

The match between the T2 sequence and the C1/C2 sequence
is

15

Seq. Id. = 172 Position = 215 to 291

AATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCTATATCCTCGAGGAG
AACTTCTAGTATATTCTGTA

20

A double stranded DNA loop of length 5.302 kilo-bases on
chromosome 7 is bounded on the left by a T1 sequence
25 whose identifier is 2840. This T1 control element has
the DNA sequence

Seq. Id. = 173 Position = 1 to 313

30 TCTGTTGGAATAAAAATCCACTATCGTCTATCAACTAATAGTTATATTATCAATATA
TTATCATATACGGTGTTAAGATGATGACATAAGTTATGAGAAGCTGTCATCGAAGTT
AGAGGAAGCTGAAACGCAAGGATTGATAATGTAATAGGATCAATGAATATAAACATA
TAAAACGGAATGAGGAATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTT

GAGGATTCCTATATCCTCGAGGAGAACTTCTAGTATATTCTGTATACCTAAATTATA
GCCTTTATCAACAATGGAATCCCAACAA

This double stranded DNA loop is bounded on the right by
5 a T2 control element whose identifier is 2859. This T2
control element has the DNA sequence

Seq. Id. = 174 Position = 1 to 314

10 CTATCAACTAATAGTTATATTATCAATATATTATCATATACGGTGTTAAGATGATGA
CATAAGTTATGAGAAGCTGTCATCGAAGTTAGAGGAAGCTGAAACGCAAGGATTGAT
AATGTAATAGGATCAATGAATATAAACATATAAACGGAATGAGGAATAATCGTAAT
ATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCTATATCCTCGAGGAGAAC
TTCTAGTATATTCTGTATACCTAATATTATAGCCTTTATCAACAATGGAATCCCAAC
15 AATTATCTCAACATTCACATATTTCTCAT

The expression of genes in this T1/T2 long loop is
controlled by the following C1/C2 short loops.

20 A C1/C2 short loop on chromosome 2 whose identifier is
298 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene YBL005W-B and has
the DNA sequence

25

Seq. Id. = 175 Position = 1 to 342

ATCTATTACATTATGGGTGGTATGTTGGAATAAAAATCCACTATCGTCTATCAACTA
ATAGTTATATTATCAATATATTATCATATACGGTGTTAAGATGATGACATAAGTTAT
30 GAGAAGCTGTCATCGAAGTTAGAGGAAGCTGAAGTGCAAGGATTGATAATGTAATAG
GATAATGAAACATATAAACGGAATGAGGAATAATCGTAATATTAGTATGTAGAAAT
ATAGATTCCATTTTGAGGATTCCTATATCCTTGAGGAGAACTTCTAGTATATTCTGT
ATACCTAATATTATAGCCTTTATCAACAATGGAATCCCAACAATTATCTCAACATTC

The match between the T1 sequence and the C1/C2 sequence is

5 Seq. Id. = 175 Position = 23 to 147

TGTTGGAATAAAAATCCACTATCGTCTATCAACTAATAGTTATATTATCAATATATT
ATCATATACGGTGTTAAGATGATGACATAAGTTATGAGAAGCTGTCATCGAAGTTAG
AGGAAGCTGAA

10

The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 175 Position = 48 to 146

15

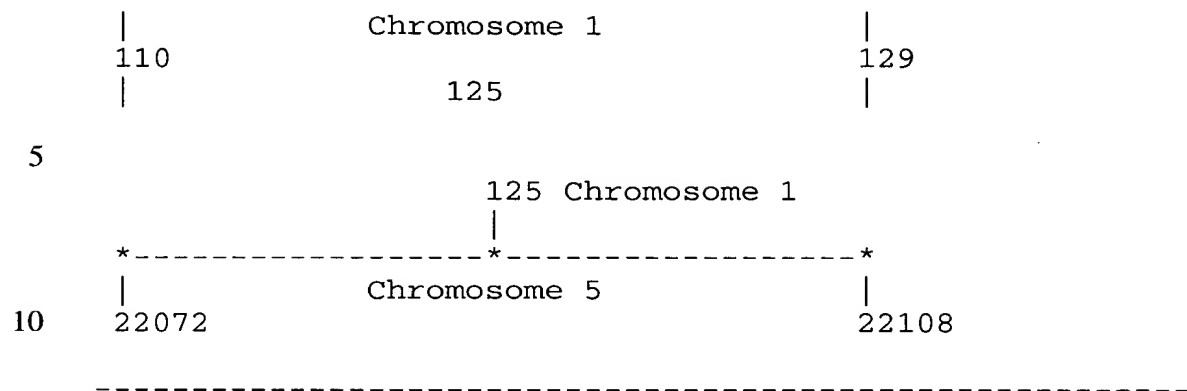
CTATCAACTAATAGTTATATTATCAATATATTATCATATACGGTGTTAAGATGATGA
CATAAGTTATGAGAAGCTGTCATCGAAGTTAGAGGAAGCTGAA

20

Example of a multi-celled transient connectron - C.
elegans

In this example the existence of the T1-T2 (22072-22108)
25 long loop is controlled by the C1/C2 (125) short loop.
The expression of this C1/C2 short loop is controlled by
the existence of the T1-T2 (110-129) long loop. The
existence of this T1-T2 long loop is itself determined by
the expression of the C1/C2 (16859) short loop. The
30 C1/C2 (125) short loop is the transient connectron.

16859 Chromosome 4
|



A double stranded DNA loop of length 18.855 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 110. This T1 control element has the DNA sequence

Seq. Id. = 176 Position = 1 to 33

20

AGCTTAGGCTTAAGCTTAGGCTTAAGCTTAGGC

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 129. This T2 control element has the DNA sequence

Seq. Id. = 177 Position = 1 to 2123

TTCTCCCGCATTTTTTGTAGATCTACGTAGATCAAACCGAAATGAGGCACTTTCTGA
ATCCACGAGCTAGGCTTAAGCTTAGGCTTAAGCTTAGGCCTTTTCTCAGGCTTAGGC
TTAGGCTTA

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

35

ZC123.3 ZC123.2

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

5 A C1/C2 short loop on chromosome 1 whose identifier is 125 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene ZC123.3 and has the DNA sequence

10 Seq. Id. = 178 Position = 1 to 89

ACGCGCCGTAAATCTACCCCAGATATGGCCGAGCCAAAATGGCCTAGTTCGGCAAAC
TCTTTCATTTCAATTTATGAGGGAAGCCAGAA

15 The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 4 whose identifier is 16859 controls the expression of the genes in this T1/T2
20 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene F58E2.7 and has the DNA sequence

Seq. Id. = 179 Position = 1 to 166

25

CTTAGGCTTAAGCTTAGGCTTAAGCTTAGGCTTAAGCTTAGGCTTAAGCTTAGGCTT
AAGCTTAGGCTTAAGCTTAGGCTTAAGCTTAGGCTTAAGCTTAGGCTTAAGCTTAGG
CTTAAGCTTAGGCTTAAGCTTAGGCTTAAGCTTAGGCTTAAGCTTAGACTTA

30 The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 179 Position = 11 to 43

AGCTTAGGCTTAAGCTTAGGCTTAAGCTTAGGC

5 The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. = 179 Position = 3 to 33

TAGGCTTAAGCTTAGGCTTAAGCTTAGGC

10

15 A double stranded DNA loop of length 51.031 kilo-bases on
chromosome 5 is bounded on the left by a T1 sequence
whose identifier is 22072. This T1 control element has
the DNA sequence

Seq. Id. = 180 Position = 1 to 57

20 CGCAACGCGCCGTAAATCTACCCCAGATATGGCCGAGCCAAAATGACCTAGTTCGGC

This double stranded DNA loop is bounded on the right by
a T2 control element whose identifier is 22108. This T2
control element has the DNA sequence

25

Seq. Id. = 181 Position = 1 to 170

TGACAATCGCCTGCCGGACAACGCGTGGAAAAGTGTCTGTACTCCACACGGACAAA
TACATTTAGTTTTTACAATAAATCGAACCGCGACGCGACACGCAACGCGACGTAAA
30 TCTACCCCAGATATGGCCGAGCCAAAATGGCCTAGTTCGGCAAACCTCTTCTATTTC

This long T1/T2 double stranded DNA loop modulates the
expression of the following genes

F36H9.3 F36H9.4 F36H9.5 F36H9.2 F36H9.1
F36H9.6

5 The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

10 A C1/C2 short loop on chromosome 1 whose identifier is 125 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene ZC123.3 and has the DNA sequence

Seq. Id. = 182 Position = 1 to 89

15

ACGCGCCGTAAATCTACCCCAGATATGGCCGAGCCAAAATGGCCTAGTTCGGCAAAC
TCTTTCATTTCAATTTATGAGGGAAGCCAGAA

20 The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 182 Position = 1 to 41

ACGCGCCGTAAATCTACCCCAGATATGGCCGAGCCAAAATG

25

The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 182 Position = 7 to 61

30

CGTAAATCTACCCCAGATATGGCCGAGCCAAAATGGCCTAGTTCGGCAAACCTTT

THIS SHEET INTENTIONALLY LEFT BLANK.

whose identifier is 1704. This T1 control element has the DNA sequence

Seq. Id. = 183 Position = 1 to 71

5

CGCCCCGTTACACGATTCCTCTGTAGTTCAGTCGGTAGAACGGCGGACTGTTAATC
CGTATGTCACTGGT

10 This double stranded DNA loop is bounded on the right by
a T2 control element whose identifier is 1718.. This T2
control element has the DNA sequence

Seq. Id. = 184 Position = 1 to 71

15 TTCAGTCGGTAGAACGGCGGACTGTTAATCCGTATGTCACTGGTTCGAGTCCAGTCA
GAGGAGCCAAATTC

This long T1/T2 double stranded DNA loop modulates the
expression of the following genes

20

asnT	b1978	b1979	b1980	shiA
amn	b1983	asnW		yeeO
asnU				

25 This long T1/T2 double stranded DNA loop modulates the
expression of the following C1/C2 short loops

30 A C1/C2 short loop on chromosome 1 whose identifier is
1705 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop is
expressed as a RNA single strand that is 3'UTR to the
gene and has the DNA sequence

Seq. Id. = 185 Position = 1 to 98

CGCCCCGTTACACGATTCCTCTGTAGTTCAGTCGGTAGAACGGCGGACTGTTAATC
CGTATGTCACTGGTTCGAGTCCAGTCAGAGGAGCCAAATTC

5

A C1/C2 short loop on chromosome 1 whose identifier is
1713 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop is
expressed as a RNA single strand that is 3'UTR to the
10 gene asnW and has the DNA sequence

Seq. Id. = 186 Position = 1 to 86

CACGATTCCTCTGTAGTTCAGTCGGTAGAACGGCGGACTGTTAATCCGTATGTCACT
15 GGTTCGAGTCCAGTCAGAGGAGCCAAATT

The expression of genes in this T1/T2 long loop is
controlled by the following C1/C2 short loops.

20 A C1/C2 short loop on chromosome 1 whose identifier is
1705 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene and has the DNA
sequence

25

Seq. Id. = 187 Position = 1 to 98

CGCCCCGTTACACGATTCCTCTGTAGTTCAGTCGGTAGAACGGCGGACTGTTAATC
CGTATGTCACTGGTTCGAGTCCAGTCAGAGGAGCCAAATTC

30

The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 187 Position = 1 to 71

CGCCCCGTTACACGATTCCTCTGTAGTTCAGTCGGTAGAACGGCGGACTGTTAATC
CGTATGTCACTGGT

5

The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. = 187 Position = 28 to 98

10

TTCAGTCGGTAGAACGGCGGACTGTTAATCCGTATGTCACTGGTTCGAGTCCAGTCA
GAGGAGCCAAATTC

15 A C1/C2 short loop on chromosome 1 whose identifier is
1713 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene asnW and has the
DNA sequence

20 Seq. Id. = 188 Position = 1 to 86

CACGATTCCTCTGTAGTTCAGTCGGTAGAACGGCGGACTGTTAATCCGTATGTCACT
GGTTCGAGTCCAGTCAGAGGAGCCAAATT

25 The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 188 Position = 1 to 60

30 CACGATTCCTCTGTAGTTCAGTCGGTAGAACGGCGGACTGTTAATCCGTATGTCACT
GGT

The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 188 Position = 17 to 86

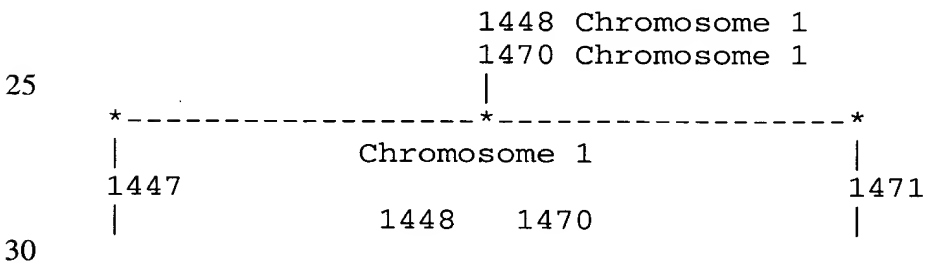
5

TTCAGTCGGTAGAACGGCGGACTGTTAATCCGTATGTCACTGGTTCGAGTCCAGTCA
GAGGAGCCAAATT

10

Example of a archea self-limiting connectrons - M.
jannaschii

In this example the existence of the T1-T2 (1447-1471)
15 long loop is controlled by two C1/C2 (1448 and 1470)
short loops. The expression of these C1/C2 short loops
is controlled by the existence of the T1-T2 (1447-1471)
long loop. The existence of this T1-T2 long loop is
itself determined by the expression of the two C1/C2
20 (1705 and 1713) short loops. The C1/C2 (1448 and 1470)
short loops are the self-limiting connectrons.



A double stranded DNA loop of length 22.675 kilo-bases on
chromosome 1 is bounded on the left by a T1 sequence
35 whose identifier is 1447. This T1 control element has
the DNA sequence

Seq. Id. = 189 Position = 1 to 95

TTATAGAACATTATGAAGCTTTTTACTCAACTAACAACCGTATCGAATTTACCATTA
5 CTTGGAAATCTATTTTAAAACCTCTTTAATCTTATGATA

This double stranded DNA loop is bounded on the right by
a T2 control element whose identifier is 1471. This T2
control element has the DNA sequence

10

Seq. Id. = 190 Position = 1 to 95

CAACTAACAACCGTATCGAATTTACCATTTACTTGGAAATCTATTTAAAACCTCTTTA
ATCTTGTGATAATAAATTCTAATCGATTCTGTGACTTAT

15

This long T1/T2 double stranded DNA loop modulates the
expression of the following genes

MJ1402	MJ1403	MJ1404	MJ1405	MJ1406
20 MJ1407	MJ1408	MJ1409	MJ1410	MJ1411
MJ1412	MJ1413	MJ1414	MJ1415	MJ1416
MJ1417	MJ1418	MJ1419	MJ1420	

This long T1/T2 double stranded DNA loop modulates the
25 expression of the following C1/C2 short loops

A C1/C2 short loop on chromosome 1 whose identifier is
1448 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop is
30 expressed as a RNA single strand that is 3'UTR to the
gene MJ1401 and has the DNA sequence

Seq. Id. = 191 Position = 1 to 122

TTATAGAACATTATGAAGCTTTTTACTCAACTAACAACCGTATCGAATTTACCATTA
CTTGGAATCTATTTAAACCTCTTTAATCTTATGATAATAAATTCTAATCGATTCTG
TGAATTAT

5

A C1/C2 short loop on chromosome 1 whose identifier is
1470 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop is
expressed as a RNA single strand that is 3'UTR to the
10 gene MJ1420 and has the DNA sequence

Seq. Id. = 192 Position = 1 to 116

TTATAGAACATTATGAAGCTTTTTACTCAACTAACAACCGTATCGAATTTACCATTA
15 CTTGGAATCTATTTAAACCTCTTTAATCTTGTGATAATAAATTCTAATCGATTCTG
TG

The expression of genes in this T1/T2 long loop is
controlled by the following C1/C2 short loops.

20

A C1/C2 short loop on chromosome 1 whose identifier is
1470 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene MJ1420 and has
25 the DNA sequence

Seq. Id. = 193 Position = 1 to 116

TTATAGAACATTATGAAGCTTTTTACTCAACTAACAACCGTATCGAATTTACCATTA
30 CTTGGAATCTATTTAAACCTCTTTAATCTTGTGATAATAAATTCTAATCGATTCTG
TG

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 193 Position = 1 to 89

5

TTATAGAACATTATGAAGCTTTTTACTCAACTAACAACCGTATCGAATTTACCATTA
CTTGGAATCTATTTAAACCTCTTTAATCTT

The match between the T2 sequence and the C1/C2 sequence is

10

Seq. Id. = 193 Position = 28 to 116

CAACTAACAACCGTATCGAATTTACCATTACTTGGAATCTATTTAAACCTCTTTA
ATCTTGTGATAATAAATTCTAATCGATTCGTG

15

A C1/C2 short loop on chromosome 1 whose identifier is 1448 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene MJ1401 and has the DNA sequence

20

Seq. Id. = 194 Position = 1 to 122

25 TTATAGAACATTATGAAGCTTTTTACTCAACTAACAACCGTATCGAATTTACCATTA
CTTGGAATCTATTTAAACCTCTTTAATCTTATGATAATAAATTCTAATCGATTG
TGACTTAT

The match between the T1 sequence and the C1/C2 sequence is

30

Seq. Id. = 194 Position = 1 to 95

TTATAGAACATTATGAAGCTTTTTACTCAACTAACAACCGTATCGAATTTACCATTA
CTTGGAATCTATTTAAACCTCTTTAATCTTATGATA

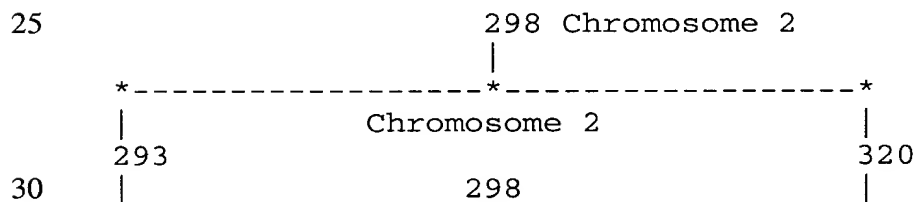
The match between the T2 sequence and the C1/C2 sequence
5 is

Seq. Id. = 194 Position = 29 to 99

CAACTAACAACCGTATCGAATTTACCATTACTTGGAAATCTATTTAAACCTCTTTA
10 ATCTT

Example of a single-celled self-limiting connectron - *S.*
15 *cervisiae*

In this example the existence of the T1-T2 (293-320)
long loop is controlled by C1/C2 (298) short loop. The
expression of this C1/C2 short loop is controlled by the
20 existence of the T1-T2 (293-320) long loop. The
existence of this T1-T2 long loop is itself determined by
the expression of the C1/C2 (298) short loop. The C1/C2
(298) short loop is the self-limiting connectron.



A double stranded DNA loop of length 38.470 kilo-bases on
35 chromosome 2 is bounded on the left by a T1 sequence

whose identifier is 293. This T1 control element has the DNA sequence

Seq. Id. = 195 Position = 1 to 258

5

GAATTGTTGGAATAAAAATCCACTATCGTCTATCAACTAATAGTTATATTATCAATA
TATTATCATATACGGTGTTAAGATGATGACATAAGTTATGAGAAGCTGTCATCGAAG
TTAGAGGAAGCTGAAGTGCAAGGATTGATAATGTAATAGGATAATGAAACATATAAA
ACGGAATGAGGAATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTTGAGG
10 ATTCCTATATCCTTGAGGAGAACTTCTAGT

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 320. This T2 control element has the DNA sequence

15

Seq. Id. = 196 Position = 1 to 77

AATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCTATATCCTCGAGGAG
AACTTCTAGTATATTCTGTA

20

This long T1/T2 double stranded DNA loop modulates the expression of the following genes

YBL005W-B	TS(AGA)B	YBL004W	YBL003C	YBL002W
25 YBL001C	YBR001C	YBR002C	YBR003W	YBR004C
YBR005W	YBR006W	YBR007C	YBR008C	YBR009C
YBR010W	YBR011C	YBR012C		

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

30

A C1/C2 short loop on chromosome 2 whose identifier is 298 controls the expression of the genes of one or more

other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YBL005W-B and has the DNA sequence

5 Seq. Id. = 5197 Position = 1 to 342

ATCTATTACATTATGGGTGGTATGTTGGAATAAAAATCCACTATCGTCTATCAACTA
ATAGTTATATTATCAATATATTATCATATACGGTGTTAAGATGATGACATAAGTTAT
GAGAAGCTGTCATCGAAGTTAGAGGAAGCTGAAGTGCAAGGATTGATAATGTAATAG
10 GATAATGAAACATATAAAACGGAATGAGGAATAATCGTAATATTAGTATGTAGAAAT
ATAGATTCCATTTTGAGGATTCCTATATCCTTGAGGAGAACTTCTAGTATATTCTGT
ATACCTAATATTATAGCCTTTATCAACAATGGAATCCCAACAATTATCTCAACATTC

The expression of genes in this T1/T2 long loop is
15 controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 2 whose identifier is
298 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
20 single strand that is 3'UTR to the gene YBL005W-B and has
the DNA sequence

Seq. Id. = 198 Position = 1 to 342

25 ATCTATTACATTATGGGTGGTATGTTGGAATAAAAATCCACTATCGTCTATCAACTA
ATAGTTATATTATCAATATATTATCATATACGGTGTTAAGATGATGACATAAGTTAT
GAGAAGCTGTCATCGAAGTTAGAGGAAGCTGAAGTGCAAGGATTGATAATGTAATAG
GATAATGAAACATATAAAACGGAATGAGGAATAATCGTAATATTAGTATGTAGAAAT
ATAGATTCCATTTTGAGGATTCCTATATCCTTGAGGAGAACTTCTAGTATATTCTGT
30 ATACCTAATATTATAGCCTTTATCAACAATGGAATCCCAACAATTATCTCAACATTC

The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. = 198 Position = 23 to 276

5 TGGTTGGAATAAAAATCCACTATCGTCTATCAACTAATAGTTATATTATCAATATATT
ATCATATACGGTGTAAAGATGATGACATAAGTTATGAGAAGCTGTCATCGAAGTTAG
AGGAAGCTGAAGTGCAAGGATTGATAATGTAATAGGATAATGAAACATATAAAACGG
AATGAGGAATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTC
CTATATCCTTGAGGAGAACTTCTAGT

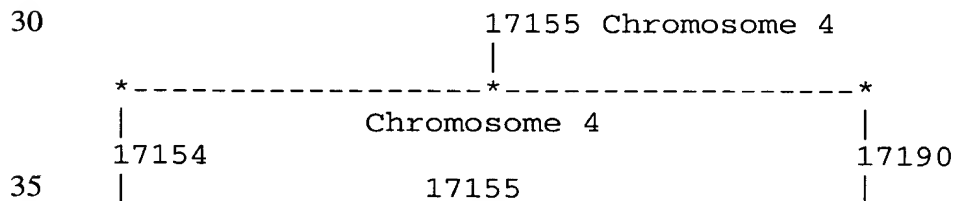
10 The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. = 198 Position = 210 to 259

15 AATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCTATATCCT

20 Example of a multi-celled self-limiting connectron - C.
elegans

25 In this example the existence of the T1-T2 (293-320)
long loop is controlled by C1/C2 (298) short loop. The
expression of this C1/C2 short loop is controlled by the
existence of the T1-T2 (293-320) long loop. The
existence of this T1-T2 long loop is itself determined by
the expression of the C1/C2 (298) short loop. The C1/C2
(298) short loop is the self-limiting connectron.



A double stranded DNA loop of length 89.919 kilo-bases on
5 chromosome 4 is bounded on the left by a T1 sequence
whose identifier is 17154. This T1 control element has
the DNA sequence

Seq. Id. = 199 Position = 1 to 29

10

AAATTTCCGGCAAATCGGCAAACCTGGCAA

This double stranded DNA loop is bounded on the right by
a T2 control element whose identifier is 17190. This T2
15 control element has the DNA sequence

Seq. Id. = 200 Position = 1 to 29

20

AATTTGCCGATTTGCCGAATTTGTCTGACA

This long T1/T2 double stranded DNA loop modulates the
expression of the following genes

R08C7.11	M01H9.2	M01H9.3	M01H9.4	M01H9.1
25 ZK180.1	ZK180.2	ZK180.3	ZK180.4	ZK180.5
ZK180.6	ZK185.3	ZK185.2		

This long T1/T2 double stranded DNA loop modulates the
expression of the following C1/C2 short loops

30

A C1/C2 short loop on chromosome 4 whose identifier is
17155 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop is

expressed as a RNA single strand that is 3'UTR to the gene R08C7.1 and has the DNA sequence

Seq. Id. = 201 Position = 1 to 56

5

AAATTTCCGGCAAATCGGCAAAC TGGCAATTTGCCGATTTGCCGAATTTGTCGACA

A C1/C2 short loop on chromosome 4 whose identifier is 17171 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene ZK180.2 and has the DNA sequence

Seq. Id. = 202 Position = 1 to 56

15

TGGAAATTTTCAGAATTTCAATTTTAATCGGCAAAATTGTACGCATCCTATGAATTT

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

20

A C1/C2 short loop on chromosome 4 whose identifier is 17155 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene R08C7.1 and has the DNA sequence

Seq. Id. = 203 Position = 1 to 56

AAATTTCCGGCAAATCGGCAAAC TGGCAATTTGCCGATTTGCCGAATTTGTCGACA

30

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 203 Position = 1 to 29

AAATTTCGGCAAATCGGCAAACCTGGCAA

- 5 The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 203 Position = 28 to 56

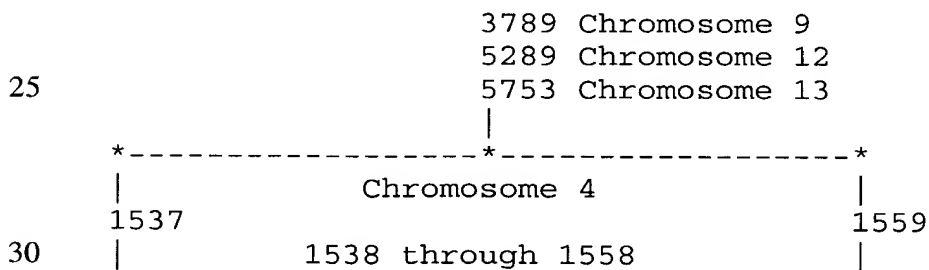
- 10 AATTGCGGATTTGCGGAATTTGTCGACA

9. Geneless connectrons exist in single-celled and multi-celled eukaryotes

5 Normally T1-T2 long loops contain genes whose expression is regulated by the existence of the long loop. When a T1-T2 long loop does not contain any genes it is described as being "geneless". The existence of the T1-T2 long loop is itself controlled by one or more C1/C2
10 short loops that may be on the same or different chromosomes. The geneless T1-T2 long loops must contain one or more C1/C2 short loops.

Example of a single-celled geneless connectron - *S. cerevisiae*
15

In this example the existence of the T1-T2 (1537-1559) long loop is controlled by three C1/C2 (3789, 5289 and 5753) short loops. The expression of 21 C1/C2 (1538 through 1558) short loops are controlled by the existence
20 of the T1-T2 (1537-1559) long loop.



A double stranded DNA loop of length 4.825 kilo-bases on
35 chromosome 4 is bounded on the left by a T1 sequence

whose identifier is 1537. This T1 control element has the DNA sequence

Seq. Id. = 204 Position = 1 to 362

5
ATGAGATATATGTGGGTAATTAGATAATTGTTGGGATTCCATTGTTGATAAAGGCTA
TAATATTAGGTATACAGAATATACTAGAAGTTCTCCTCGAGGATTTAGGAATCCATA
AAAGGGAATCTGCAATTCTACACAATTCTATAAATATTATTATCATCGTTTTATATG
TTAATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCAGCTTCCACTAATT
10 TAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACACCGTATATGATAATATAC
TAGTAACGTAAATACTAGTTAGTAGATGATAGTTGATTTTTATTCCAACATACCACC
CATAATGTAATAGATCTAAT

This double stranded DNA loop is bounded on the right by
15 a T2 control element whose identifier is 1559. This T2
control element has the DNA sequence

Seq. Id. = 205 Position = 1 to 362

20 ATGAGATATATGTGGGTAATTAGATAATTGTTGGGATTCCATTGTTGATAAAGGCTA
TAATATTAGGTATACAGAATATACTAGAAGTTCTCCTCGAGGATTTAGGAATCCATA
AAAGGGAATCTGCAATTCTACACAATTCTATAAATATTATTATCATCGTTTTATATG
TTAATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCAGCTTCCACTAATT
TAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACACCGTATATGATAATATAC
25 TAGTAACGTAAATACTAGTTAGTAGATGATAGTTGATTTTTATTCCAACATACCACC
CATAATGTAATAGATCTAAT

There are no genes controlled by this T1/T2 loop.

30 This long T1/T2 double stranded DNA loop modulates the
expression of the following C1/C2 short loops

A C1/C2 short loop on chromosome 4 whose identifier is 1538 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

5

Seq. Id. = 206 Position = 1 to 387

ATGAGATATATGTGGGTAATTAGATAATTGTTGGGATTCCATTGTTGATAAAGGCTA
TAATATTAGGTATACAGAATATACTAGAAGTTCTCCTCGAGGATTTAGGAATCCATA
10 AAAGGGAATCTGCAATTCTACACAATTCTATAAATATTATTATCATCGTTTTATATG
TTAATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCAGCTTCCACTAATT
TAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACACCGTATATGATAATATAC
TAGTAACGTAAATACTAGTTAGTAGATGATAGTTGATTTTTATTCCAACATACCACC
CATAATGTAATAGATCTAATGAATCCATTTGTTTGTTAATAGTTT

15

This T1-T2 loop also modulates the C1/C2 short loops numbered 1539 to 1557

A C1/C2 short loop on chromosome 4 whose identifier is 1558 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

20 Seq. Id. = 207 Position = 1 to 307

25

AGCTTCTCATAACTTATGTCATCATCTTAACACCGTATATGATAATATATTGATAAT
ATAACTTGTTGGAATAAAAATCAACTATCATCTACTAAGTAGTATTTACGTTACTAG
TATATTATCATATACGGTGTTAGAAGATGACGCAAATGATGAGAAATAGTCATCTAA
ATTAGTGGAAGCTGA...GTCTATCTGGCGAATATAAATTTTTACGCTACACACGTC
30 ATCGACATCTAAATATGACAGTCGCTGAACTGTTCTTAGATATCCATGCTATTTATG
AAGAACAACAGGGATCGAGAAACAG

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

5 A C1/C2 short loop on chromosome 9 whose identifier is 3789 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YIL059C and has the DNA sequence

10 Seq. Id. = 208 Position = 1 to 176

TTTATATGTTAATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCAGCTTC
CACTAATTTAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACACCGTATATGA
TAATATACTAGTAACGTAAATACTAGTTAGTAGATGATAGTTGATTTTTATTCCAAC
15 AGTAT

The match between the T1 sequence and the C1/C2 sequence is

20 Seq. Id. = 208 Position = 1 to 172

TTTATATGTTAATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCAGCTTC
CACTAATTTAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACACCGTATATGA
TAATATACTAGTAACGTAAATACTAGTTAGTAGATGATAGTTGATTTTTATTCCAAC
25 A

The match between the T2 sequence and the C1/C2 sequence is

30 Seq. Id. = 208 Position = 1 to 172

TTTATATGTTAATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCAGCTTC
CACTAATTTAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACACCGTATATGA

TAATATACTAGTAACGTAAATACTAGTTAGTAGATGATAGTTGATTTTTATTCCAAC
A

5 A C1/C2 short loop on chromosome 12 whose identifier is
5289 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene YLR301W and has
the DNA sequence

10 Seq. Id. =209 Position = 1 to 325

GGTGAATTTTGAGATAATTGTTGGGATTCCATTTTAAATAAGGCAATAATATTAGGT
ATGTAGAATATACTAGAAGTTCTCCTCGAGGATTTAGGAATCCATAAAAGGGAATCT
GCAATTCTACACAATTCTATAAATATTATTATCATCGTTTTATATGTTAATATTCAT
15 TGATCCTATTACATTATCAATCCTTGCGTTTCAGCTTCCACTAATTTAGATGACTAT
TTCTCATCATTTGCGTCATCTTCTAACACCGTATATGATAATATACTAGTAACGTAA
ATACTAGTTAGTAGATGATAGTTGATTTTTATTCCAACAC

20 The match between the T1 sequence and the C1/C2 sequence
is

Seq. Id. =209 Position = 62 to 317

25 AGAATATACTAGAAGTTCTCCTCGAGGATTTAGGAATCCATAAAAGGGAATCTGCAA
TTCTACACAATTCTATAAATATTATTATCATCGTTTTATATGTTAATATTCATTGAT
CCTATTACATTATCAATCCTTGCGTTTCAGCTTCCACTAATTTAGATGACTATTTCT
CATCATTTGCGTCATCTTCTAACACCGTATATGATAATATACTAGTAACGTAAATAC
TAGTTAGTAGATGATAGTTGATTTTTATTCCAACA

30 The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. =209 Position = 86 to 324

AGGATTTAGGAATCCATAAAAGGGAATCTGCAATTCTACACAATTCTATAAATATTA
TTATCATCGTTTTATATGTTAATATTCATTGATCCTATTACATTATCAATCCTTGCG
TTTCAGCTTCCACTAATTTAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACA
5 CCGTATATGATAATATACTAGTAACGTAAATACTAGTTAGTAGATGATAGTTGATTT
TTATTCCAACA

A C1/C2 short loop on chromosome 13 whose identifier is
5753 controls the expression of the genes in this T1/T2
10 long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene YMR044W and has
the DNA sequence

Seq. Id. = 210 Position = 1 to 334

15 TTGAGAAATGGGGGAATGTTGAGATAATTGTTGGGATTCCATTGTTGATAAAGGCTA
TAATATTAGGTATACAGAATATACTAGAAGTTCTCCTCAAGGATATAGGAATCCTCA
AAATGGAATCTATATTTCTACATACTAATATTACGATTATTCCTCATTCGGTTTTAT
ATGTTTTCATTATCCTATTACATTATCAATCCTTGCACTTCAGCTTCCTCTAACTTCG
20 ATGACAGCTTCTCATAACTTATGTCATCATCTTAACACCGTATATGATAATATATTG
ATAATATAACTATTAGTTGATAGACGATAGTGGATTTTTATTCCAACAT

The match between the T1 sequence and the C1/C2 sequence
is

25 Seq. Id. = 210 Position = 22 to 95

AGATAATTGTTGGGATTCCATTGTTGATAAAGGCTATAATATTAGGTATACAGAATA
TACTAGAAGTTCTCCTC

30 The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. = 210 Position = 28 to 101

TTGTTGGGATTCCATTGTTGATAAAGGCTATAATATTAGGTATACAGAATATACTAG
AAGTTCTCCTCAAGGAT

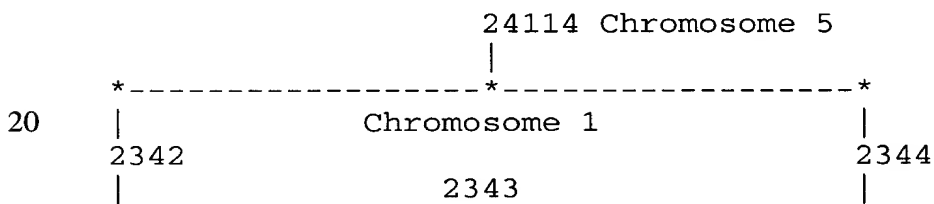
5

Two examples of multi-celled geneless connectrons - *C. elegans*

10

In the first example the existence of the T1-T2 (2342-2344) long loop is controlled by the C1/C2 (24114) short loop. The expression of one C1/C2 (2343) short loop is controlled by the existence of the T1-T2 (2342-2344) long loop.

15

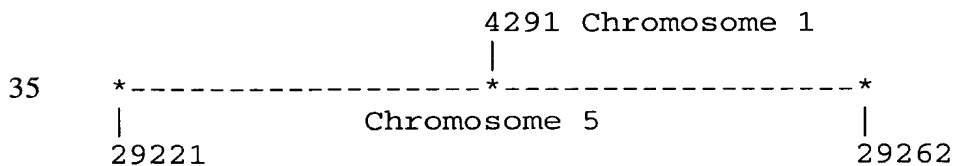


20

25

In the second example the existence of the T1-T2 (29221-29262) long loop is controlled by the C1/C2 (4291) short loop. The expression of one C1/C2 (2343) short loop is controlled by the existence of the T1-T2 (2342-2344) long loop.

30



35

| 29222 through 29261 |

5 A double stranded DNA loop of length 67.059 kilo-bases on chromosome 1 is bounded on the left by a T1 sequence whose identifier is 2342. This T1 control element has the DNA sequence

10 Seq. Id. = 211 Position = 1 to 37

TGAAAACTACAGTAATTCTTTAAATGACTACTGTAGC

This double stranded DNA loop is bounded on the right by
15 a T2 control element whose identifier is 2344. This T2 control element has the DNA sequence

Seq. Id. = 212 Position = 1 to 37

20 CTACTGTAGCGCTTGTGTCGATTTACGGGCTCGATTT

There are no genes controlled by this T1/T2 loop.

This long T1/T2 double stranded DNA loop modulates the
25 expression of the following C1/C2 short loops

A C1/C2 short loop on chromosome 1 whose identifier is 2343 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the
30 DNA sequence

Seq. Id. = 213 Position = 1 to 61

TCGACACAAGCGCTACAGTAGCTATTTAAAGAATTACTGTAGTTTTTCGCTACGAGAT
ATTT

The expression of genes in this T1/T2 long loop is
5 controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 5 whose identifier is
24114 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
10 single strand that is 3'UTR to the gene C13F10.5 and has
the DNA sequence

Seq. Id. = 214 Position = 1 to 68

15 GCGAAACTACAGTAATTCTTTAAATGACTACTGTAGCGCTTGTGTCGATTTACGGG
CTCGATTTTCG

The match between the T1 sequence and the C1/C2 sequence
is

20

Seq. Id. = 214 Position = 3 to 38

GAAACTACAGTAATTCTTTAAATGACTACTGTAGC

25 The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. = 214 Position = 29 to 65

30 C TACTGTAGCGCTTGTGTCGATTTACGGGCTCGATT

A double stranded DNA loop of length 41.297 kilo-bases on chromosome 5 is bounded on the left by a T1 sequence whose identifier is 29221. This T1 control element has the DNA sequence

5

Seq. Id. = 215 Position = 1 to 62

TTTAAATTTCCCGCCAAAAATTGACTGAAAATTTGGATTTTCTTTCCAAAAATTGAC
AGAAA

10

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 29262. This T2 control element has the DNA sequence

15 Seq. Id. = 216 Position = 1 to 31

TGAAAATTTGAATTTCCCGCCAAAAATTAAC

There are no genes controlled by this T1/T2 loop.

20

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

25 A C1/C2 short loop on chromosome 5 whose identifier is 29222 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

Seq. Id. = 217 Position = 1 to 58

30

AATTTCCCGCCAAAAATTGACTGAAAATTTGGATTTTCTTTCCAAAAATTGACAGAA
A

This T1-T2 loop also modulates the C1/C2 short loops numbered 29223 to 29260

5 A C1/C2 short loop on chromosome 5 whose identifier is 29261 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

Seq. Id. = 218 Position = 1 to 54

10

AAAATTGACTGAAAATTTGAATTTCCAGCCAAAAATTGACTGAAAATTTGAATT

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

15

A C1/C2 short loop on chromosome 1 whose identifier is 4291 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene Y43F8C.5 and has
20 the DNA sequence

Seq. Id. = 219 Position = 1 to 317

25 AAAATTAAGTAAAATTTGAATTTCCCGCCAAAAATTGACTGAAAATTTGAATTTCC
CGCCAAAAAAATTGACTGAAAATTTGAATTTCCCGCCAAAAATTGACTGAAAATTT
GAATTTCCCGCCAAAAATTAATTGAAAATTTGAATTTCCCGCCAAAAATTAATTGAA
ACTTTGAATTTTCAA...ATTTCCCGCCAAAAATTAATTGAACTTTGAATTTTCAA
ATTTCCCGCCAAAAATTGACTGAAAATTTGAATTTCCCGCCAAAAATTAATTGAAAA
TTTGAATTTTGAATTTCCCGCCAAAAATGACTGA

30

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 219 Position = 229 to 260

AAATTTCCCGCCAAAAATTGACTGAAAATTTG

5 The match between the T2 sequence and the C1/C2 sequence
is

Seq. Id. = 219 Position = 63 to 104

10 AAAAAAATTGACTGAAAATTTGAATTTCCCGCCAAAAATTGA

10. One connectron controls many geneless connectrons in single-celled and multi-celled eukaryotes

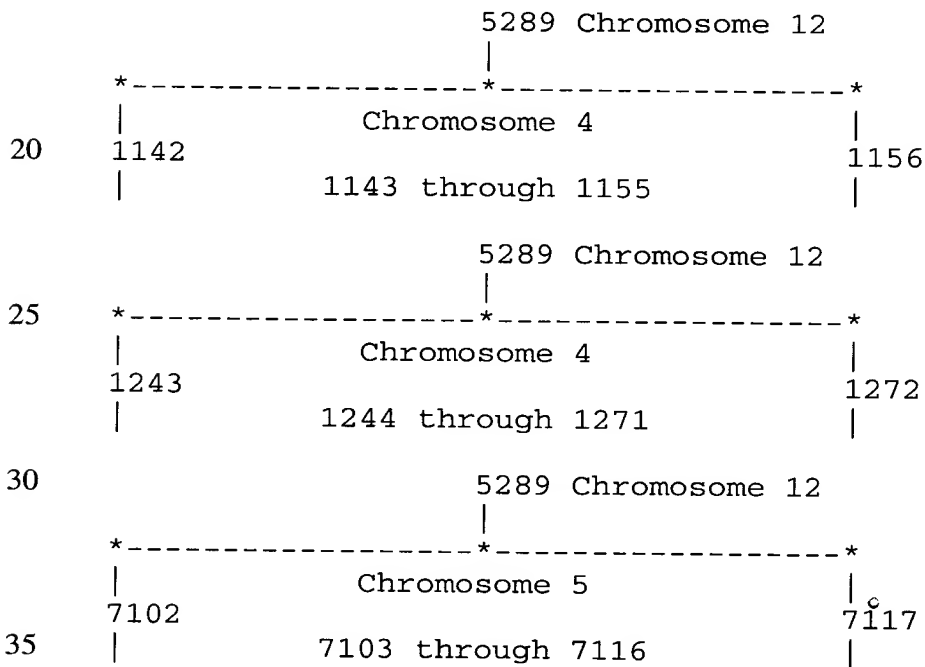
5

One C1/C2 short loop can control the existence of many geneless T1-T2 long loops.

10 Example of a single-celled geneless connectron - *S. cerevisiae*

In this example the existence of the three T1-T2 (1142-1156, 1242-1272 and 7102-7117) long loops is controlled by the C1/C2 (5289) short loop.

15



40 A double stranded DNA loop of length 5.337 kilo-bases on chromosome 4 is bounded on the left by a T1 sequence

whose identifier is 1142. This T1 control element has the DNA sequence

Seq. Id. = 220 Position = 1 to 318

5

ATTTTGAGATAATTGTTGGGATTCCATTTTAAATAAGGCAATAATATTAGGTATGTA
GATATACTAGAAGTTCTCCTCGAGGATTTAGGAATCCATAAAAGGGAATCTGCAATT
CTACACAATTCTATAAATATTATTATCATCATTTTATATGTTAATATTCATTGATCC
TATTACATTATCAATCCTTGCGTTTCAGCTTCCACTAATTTAGATGACTATTTCTCA
10 TCATTTGCGTCATCTTCTAACACCGTATATGATAATATACTAGTAACGTAAATACTA
GTTAGTAGATGATAGTTGATTTTTATTCCAACA

This double stranded DNA loop is bounded on the right by
a T2 control element whose identifier is 1156. This T2
15 control element has the DNA sequence

Seq. Id. = 221 Position = 1 to 295

TTTTAATAAGGCAATAATATTAGGTATGTAGATATACTAGAAGTTCTCCTCCAGGAT
20 TTAGGAATCCATAAAAGGGAATCTGCAATTCTACACAATTCTATAAATATTATTATC
ATCATTTTATATGTTAATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCA
GCTTCCACTAATTTAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACACCGTA
TATGATAATATACTAGTAACGTAAATACTAGTTAGTAGATGATAGTTGATTTTTATT
CCAACAAGAA

25

There are no genes controlled by this T1/T2 loop.

This long T1/T2 double stranded DNA loop modulates the
expression of the following C1/C2 short loops

30

A C1/C2 short loop on chromosome 4 whose identifier is
1143 controls the expression of the genes of one or more

other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

Seq. Id. = 222 Position = 1 to 349

5

ATTTTGAGATAATTGTTGGGATTCCATTTTAAATAAGGCAATAATATTAGGTATGTA
GATATACTAGAAGTTCTCCTCGAGGATTTAGGAATCCATAAAAGGGAATCTGCAATT
CTACACAATTCTATAAATATTATTATCATCATTTTATATGTTAATATTCATTGATCC
TATTACATTATCAAT...CTCTAAGTCTCATTGCCTTTGTGCCAAAAAATCTGTTTC
10 TAAATTTCTCTTCATTTGTAGACTTAATTATACTGATCGTTGATCTACTATCAGTAA
GTAAGCCTTTAATAATTGGTTTCTTGTTAAGTTCTTGCACAAGGTGACTGAGGTTAT
TCAATAGCGG

This T1-T2 loop also modulates the C1/C2 short loops
15 numbered 1144 to 1154

A C1/C2 short loop on chromosome 4 whose identifier is
1155 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop has the
20 DNA sequence

Seq. Id. = 223 Position = 1 to 69

GAGGAGAACTTCTAGTATATCTACATACCTAATATTATTGCCTTATTAAAAATGGAA
25 TCCCAACAATTA

The expression of genes in this T1/T2 long loop is
controlled by the following C1/C2 short loops.

30 A C1/C2 short loop on chromosome 12 whose identifier is
5289 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA

single strand that is 3'UTR to the gene YLR301W and has the DNA sequence

Seq. Id. = 224 Position = 1 to 324

5

GGTGAATTTTGAGATAATTGTTGGGATTCCATTTTAAATAAGGCAATAATATTAGGT
ATGTAGAATATACTAGAAGTTCTCCTCGAGGATTTAGGAATCCATAAAAGGGAATCT
GCAATTCTACACAATTCTATAAATATTATTATCATCGTTTTATATGTTAATATTCAT
TGATCCTATTACATTATCAATCCTTGCGTTTCAGCTTCCACTAATTTAGATGACTAT
10 TTCTCATCATTTGCGTCATCTTCTAACACCGTATATGATAATATACTAGTACGTAA
TACTAGTTAGTAGATGATAGTTGATTTTTATTCCAACAC

The match between the T1 sequence and the C1/C2 sequence is

15

Seq. Id. = 224 Position = 6 to 64

ATTTTGAGATAATTGTTGGGATTCCATTTTAAATAAGGCAATAATATTAGGTATGTA
GA

20

The match between the T2 sequence and the C1/C2 sequence is

Seq. Id. = 224 Position = 33 to 64

25

TTTTAATAAGGCAATAATATTAGGTATGTAGA

30 A double stranded DNA loop of length 5.251 kilo-bases on chromosome 4 is bounded on the left by a T1 sequence whose identifier is 1243. This T1 control element has the DNA sequence

Seq. Id. = 225 Position = 1 to 366

CGTGTTTTATCTCATGTTGTTTCGTTTTGTTATTGAGATATATGTGGGTAATTAGATA
5 ATTGTTGGGATTCCATTGTTGATAAAGGCTATAATATTAGGTATACAGAATATACTA
GAAGTTCTCCTCGAGGATTTAGGAATCCATAAAAGGGAATCTGCAATTCTACACAAT
TCTATAAATATTATTATCATCGTTTTATATGTTAATATTCATTGATCCTATTACATT
ATCAATCCTTGCGTTTCAGCTTCCACTAATTTAGATGACTATTTCTCATCATTTGCG
TCATCTTCTAACACCGTATATGATAATATACTAGTAACGTAAATACTAGTTAGTAGA
10 TGATAGTTGATTTTTATTCCAACA

This double stranded DNA loop is bounded on the right by
a T2 control element whose identifier is 1272. This T2
control element has the DNA sequence

15

Seq. Id. = 226 Position = 1 to 273

TGAGATATATGTGGGTAATTAGATAAATTGTTGGGATTCCATTGTTGATAAAGGCTAT
AATATTAGGTATACAGAATATACTAGAAGTTCTCCTCGAGGATTTAGGAATCCATAA
20 AAGGGAATCTGCAATTCTACACAATTCTATAAATATTATTATCATCGTTTTATATGT
TAATATTCATTGATC...TATACTAGTAACGTAAATACTAGTTAGTAGATGATAGTT
GATTTTTATTCCAACAGTTATAAGGTTGTTTCATATGTGTTTTATGAA

There are no genes controlled by this T1/T2 loop.

25

This long T1/T2 double stranded DNA loop modulates the
expression of the following C1/C2 short loops

A C1/C2 short loop on chromosome 4 whose identifier is
30 1244 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop has the
DNA sequence

Seq. Id. = 227 Position = 1 to 327

TTTATCTCATGTTGTTTCGTTTTGTTATTGAGATATATGTGGGTAATTAGATAATTGT
TGGGATTCCATTGTTGATAAAGGCTATAATATTAGGTATACAGAATATACTAGAAGT
5 TCTCCTCGAGGATTTAGGAATCCATAAAAGGGAATCTGCAATTCTACACAATTCTAT
AAATATTATTATCAT...GTCTCGATGTAGTATACGTATAAATTATTACCTGATACT
TCATCTCTAAGTCTCATTGCCTTTGTGCCAAAAAATCTGTTTCTAAATTTCTCTTCA
TTTGTAGACTTAATTATACTGATCGTTGATCTACTATCAGTAAGT

10 This T1-T2 loop also modulates the C1/C2 short loops
numbered 1245 to 1270

A C1/C2 short loop on chromosome 4 whose identifier is
1271 controls the expression of the genes of one or more
15 other T1/T2 long loops. This C1/C2 short loop has the
DNA sequence

Seq. Id. = 228 Position = 1 to 309

20 TGTTGTATCTCAAAATGAGATATGTCAGTATGACAATACGTCATCCTAAACGTTTCAT
AAAACACATATGAAACAACCTTATAACTGTTGGAATAAAAATCAACTATCATCTACT
AACTAGTATTTACGTTACTAGTATATTATCATATACGGTGTTAGAAGATGACGCAAA
TGATGAGAAATAGTC...CAACAATGGAATCCCAACAATTATCTAATTACCCACATA
TATCTCATGGTAGCGCCTGTGCTTCGGTACTTCTAAGGAAGTCCACACAAATCAAG
25 ATCCGTTAGACGTTTCAGCTTCCAAAA

The expression of genes in this T1/T2 long loop is
controlled by the following C1/C2 short loops.

30 A C1/C2 short loop on chromosome 12 whose identifier is
5289 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA

single strand that is 3'UTR to the gene YLR301W and has the DNA sequence

Seq. Id. = 229 Position = 1 to 325

5

GGTGAATTTTGAGATAATTGTTGGGATTCCATTTTAAATAAGGCAATAATATTAGGT
ATGTAGAATATACTAGAAAGTTCTCCTCGAGGATTTAGGAATCCATAAAAGGGAATCT
GCAATTCTACACAATTCTATAAATATTATTATCATCGTTTTATATGTTAATATTCAT
TGATCCTATTACATTATCAATCCTTGCGTTTCAGCTTCCACTAATTTAGATGACTAT
10 TTCTCATCATTTGCGTCATCTTCTAACACCGTATATGATAATATACTAGTAACGTAA
ATACTAGTTAGTAGATGATAGTTGATTTTTATTCCAACAC

The match between the T1 sequence and the C1/C2 sequence is

15

Seq. Id. = 229 Position = 62 to 317

AGAATATACTAGAAAGTTCTCCTCGAGGATTTAGGAATCCATAAAAGGGAATCTGCAA
TTCTACACAATTCTATAAATATTATTATCATCGTTTTATATGTTAATATTCATTGAT
20 CCTATTACATTATCAATCCTTGCGTTTCAGCTTCCACTAATTTAGATGACTATTTCT
CATCATTTGCGTCATCTTCTAACACCGTATATGATAATATACTAGTAACGTAAATAC
TAGTTAGTAGATGATAGTTGATTTTTATTCCAACA

The match between the T2 sequence and the C1/C2 sequence is

25

Seq. Id. = 229 Position = 62 to 317

AGAATATACTAGAAAGTTCTCCTCGAGGATTTAGGAATCCATAAAAGGGAATCTGCAA
30 TTCTACACAATTCTATAAATATTATTATCATCGTTTTATATGTTAATATTCATTGAT
CCTATTACATTATCAATCCTTGCGTTTCAGCTTCCACTAATTTAGATGACTATTTCT
CATCATTTGCGTCATCTTCTAACACCGTATATGATAATATACTAGTAACGTAAATAC
TAGTTAGTAGATGATAGTTGATTTTTATTCCAACA

A double stranded DNA loop of length 5.296 kilo-bases on
5 chromosome 15 is bounded on the left
by a T1 sequence whose identifier is 7102. This T1
control element has the DNA sequence

Seq. Id. = 230 Position = 1 to 365

10

CATGATTAATATGACCAATCGGCGTGTGTTTTTGAAAAGTGGGTGAATTTTGAGATA
ATTGTTGGGATTCCATTTTAAATAAGGCAATAATATTAGGTATGTAGAATGTACTAG
AAGTTCTCCTCAAGGATTTAGGAATCCATGAAAGGGAATCTGCAATTCTACACAATT
CTATAAATATTATTATCATCATTTTATATGTTAATATTCATTGATCCTATTACATTA
15 TCAATCCTTGCGTTTCAGCTTCCACTAATTTAGATGACTATTTCTCATCATTTGCGT
CATCTTCTAACACCGTATATGATAATATACTAGTAACGTAAATACTAGTTAGTAGAT
GATAGTTGATTTTTTATTCCAACA

This double stranded DNA loop is bounded on the right by
20 a T2 control element whose identifier is 7117. This T2
control element has the DNA sequence

Seq. Id. = 231 Position = 1 to 365

25 TGAAAAGTGGGTGAATTTTGAGATAATTGTTGGGATTCCATTTTAAATAAGGCAATA
ATATTAGGTATGTAGAATGTACTAGAAGTTCTCCTCAAGGATTTAGGAATCCATGAA
AGGGAATCTGCAATTCTACACAATTCTATAAATATTATTATCATCATTTTATATGTT
AATATTCATTGATCCTATTACATTATCAATCCTTGCGTTTCAGCTTCCACTAATTTA
GATGACTATTTCTCATCATTTGCGTCATCTTCTAACACCGTATATGATAATATACTA
30 GTAACGTAAATACTAGTTAGTAGATGATAGTTGATTTTTTATTCCAACAGTTTATAT
ACCTCTCTTATTTAGTATAAGAA

There are no genes controlled by this T1/T2 loop.

This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

5 A C1/C2 short loop on chromosome 15 whose identifier is 7103 controls the expression of the genes of one or more other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

10 Seq. Id. = 232 Position = 1 to 357

AAGAACATTGCTGATGTGATGACAAAACCTCTTCCGATAAAAACATTTAAACTATTA
ACTAACAAATGGATTTCATTAGATCTATTACATTATGGGTGGTATGTTGGAATAAAAA
TCAACTATCATCTACTAACTAGTATTTACGTTACTAGTATATTATCATATACGGTGT
15 TAGAAGATGACGCAAATGATGAGAAATAGTCATCTAAATTAGTGGAAGCTGAAACGC
AAGGATTGATAATGTAATAGGATCAATGAATATTAACATATAAAATGATGATAATAA
TATTTATAGAATTGTGTAGAATTGCAGATTCCCTTTTCATGGATTCCCTAAATCCTTGA
GGAGAACTTCTAGTA

20 This T1-T2 loop also modulates the C1/C2 short loops numbered 7104 to 7115

A C1/C2 short loop on chromosome 15 whose identifier is 7116 controls the expression of the genes of one or more
25 other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

Seq. Id. = 233 Position = 1 to 66

30 CCATTCTGTGGAGGTGGTACTGAAGCAGGTTGAGGAGAGACATGATGATGGTTCTCT
GGAACAGCT

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

5 A C1/C2 short loop on chromosome 12 whose identifier is 5289 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene YLR301W and has the DNA sequence

10 Seq. Id. = 234 Position = 1 to 325

GGTGAATTTTGAGATAATTGTTGGGATTCCATTTTAAATAAGGCAATAATATTAGGT
ATGTAGAATATACTAGAAGTTCTCCTCGAGGATTTAGGAATCCATAAAAGGGAATCT
GCAATTCTACACAATTCTATAAATATTATTATCATCGTTTTATATGTTAATATTCAT
15 TGATCCTATTACATTATCAATCCTTGCGTTTCAGCTTCCACTAATTTAGATGACTAT
TTCTCATCATTTGCGTCATCTTCTAACACCGTATATGATAATATACTAGTAACGTAA
ATACTAGTTAGTAGATGATAGTTGATTTTTATTCCAACAC

The match between the T1 sequence and the C1/C2 sequence
20 is

Seq. Id. = 234 Position = 1 to 66

GGTGAATTTTGAGATAATTGTTGGGATTCCATTTTAAATAAGGCAATAATATTAGGT
25 ATGTAGAAT

The match between the T2 sequence and the C1/C2 sequence
is

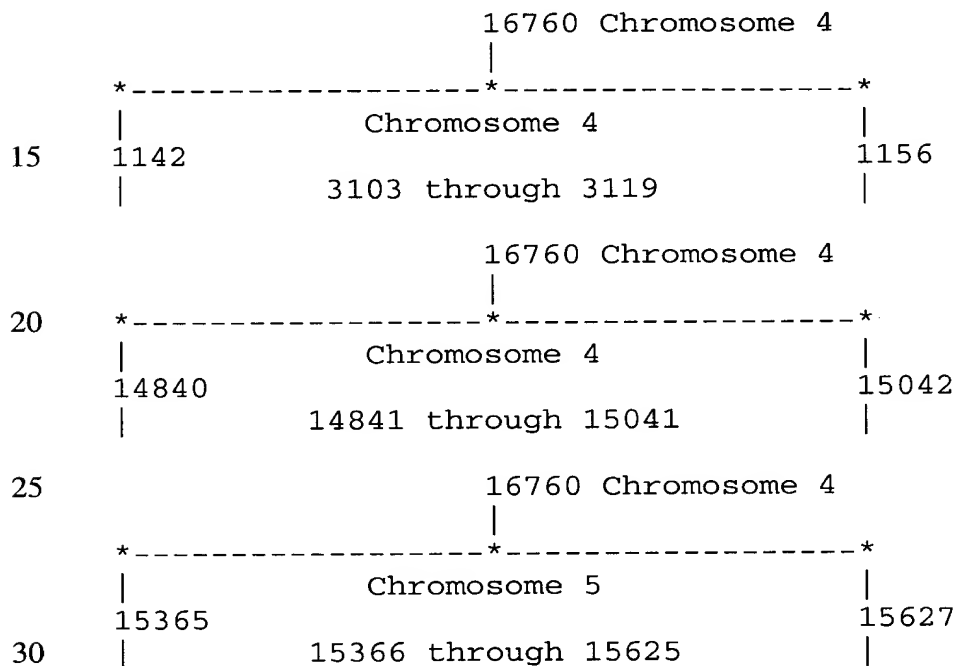
30 Seq. Id. = 234 Position = 1 to 66

GGTGAATTTTGAGATAATTGTTGGGATTCCATTTTAAATAAGGCAATAATATTAGGT
ATGTAGAAT

Example of a multi-celled geneless connectron - C.
 5 elegans

In this example the existence of the three T1-T2 (1142-1156, 14840-15042 and 15365-15627) long loops is controlled by the C1/C2 (16760) short loop.

10



A double stranded DNA loop of length 15.894 kilo-bases on
 35 chromosome 1 is bounded on the left by a T1 sequence
 whose identifier is 3101. This T1 control element has
 the DNA sequence

Seq. Id. = 235 Position = 1 to 33

40

CAAATCGGCAAATTGCCGGAATTGAACATTTC

This double stranded DNA loop is bounded on the right by
a T2 control element whose identifier is 3120. This T2
5 control element has the DNA sequence

Seq. Id. = 236 Position = 1 to 54

10 AAACGATTTTTCGGCAAATCGGCAAATTGCCGGAATTGTAATTTCCGGCAAAT

There are no genes controlled by this T1/T2 loop.

This long T1/T2 double stranded DNA loop modulates the
expression of the following C1/C2 short loops
15

A C1/C2 short loop on chromosome 1 whose identifier is
3103 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop has the
DNA sequence

20 Seq. Id. = 237 Position = 1 to 55

TTAAAATTTCCGGCAAATCGGCAAATTGGCAGAAATGAAACTCACGGCAAATCGG

25 This T1-T2 loop also modulates the C1/C2 short loops
numbered 3104 to 3118

A C1/C2 short loop on chromosome 1 whose identifier is
3119 controls the expression of the genes of one or more
30 other T1/T2 long loops. This C1/C2 short loop has the
DNA sequence

Seq. Id. = 238 Position = 1 to 61

CCCGCATTTTTTTGTAGATCAAACCGTAATGGGACGGCCTGGCAACACGTGATTTTCC
AAAT

- 5 The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

A C1/C2 short loop on chromosome 4 whose identifier is 16760 controls the expression of the genes in this T1/T2
10 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene T23E1.2 and has the DNA sequence

Seq. Id. = 239 Position = 1 to 124

15

GGCAAATTGCCGAAATTGAACATTTCCGGCAAATCGGCAAATTGCCGGAATTGAACA
TTTCCGGCAAATCGGCAAATTGCCGGAATTGAACATTTCCGGCAAATCGGCAAATTG
CCGGAATTGA

- 20 The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 239 Position = 30 to 62

- 25 CAAATCGGCAAATTGCCGGAATTGAACATTTCC

The match between the T2 sequence and the C1/C2 sequence is

- 30 Seq. Id. = 239 Position = 23 to 53

TTTCCGGCAAATCGGCAAATTGCCGGAATTG

A double stranded DNA loop of length 86.977 kilo-bases on chromosome 3 is bounded on the left by a T1 sequence
5 whose identifier is 14840. This T1 control element has the DNA sequence

Seq. Id. = 240 Position = 1 to 141

10 AAAAATTTCCGGCAAGTCGGCAATTTTCCGAAAATGAAAATTTCCGGCAAATCGGCA
AATTGCCGGAATTGAAAATTCCTGGCAAATCAGCAAATTTGCGGCAAATCGGCAATT
TGCCGAAAATGAAAATTTCCGGCAAAT

This double stranded DNA loop is bounded on the right by
15 a T2 control element whose identifier is 15042. This T2 control element has the DNA sequence

Seq. Id. = 241 Position = 1 to 98

20 CAAATCGGTAGGTAAATTGGCCAAACTTGAAAATTTCCGGCAAATCGGCAAATTCCG
CGAACTGAACATTTCCGGCAAATCGGCAAATTGCTCGAACT

There are no genes controlled by this T1/T2 loop.

25 This long T1/T2 double stranded DNA loop modulates the expression of the following C1/C2 short loops

A C1/C2 short loop on chromosome 3 whose identifier is 14841 controls the expression of the genes of one or more
30 other T1/T2 long loops. This C1/C2 short loop has the DNA sequence

Seq. Id. = 242 Position = 1 to 141

AAAAATTTCCGGCAAGTCGGCAATTTTCCGAAAATGAAAATTTCCGGCAAATCGGCA
AATTGCCGGAATTGAAAATTCCTGGCAAATCAGCAAATTTGCGGCAAATCGGCAATT
TGCCGAAAATGAAAATTTCCGGCAAAT

5

This T1-T2 loop also modulates the C1/C2 short loops
numbered 14842 to 15040

10 A C1/C2 short loop on chromosome 3 whose identifier is
15041 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop has the
DNA sequence

Seq. Id. = 243 Position = 1 to 55

15

CGGCAATTGCCGTTCCGGCAATTTGCCAATTTGCCGGAATTTTCAATTCCGGCAA

The expression of genes in this T1/T2 long loop is
controlled by the following C1/C2 short loops.

20

A C1/C2 short loop on chromosome 4 whose identifier is
16760 controls the expression of the genes in this T1/T2
long loop. This C1/C2 short loop is expressed as a RNA
single strand that is 3'UTR to the gene T23E1.2 and has
25 the DNA sequence

Seq. Id. = 244 Position = 1 to 124

30 GGCAAATTGCCGAAATTGAACATTTCCGGCAAATCGGCAAATTGCCGGAATTGAACA
TTTCCGGCAAATCGGCAAATTGCCGGAATTGAACATTTCCGGCAAATCGGCAAATTG
CCGGAATTGA

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 244 Position = 22 to 55

5

ATTTCGGCAAATCGGCAAATTGCCGGAATTGAA

The match between the T2 sequence and the C1/C2 sequence is

10

Seq. Id. = 244 Position = 17 to 45

TGAACATTTCCGGCAAATCGGCAAATTGC

15

A double stranded DNA loop of length 98.488 kilo-bases on chromosome 3 is bounded on the left by a T1 sequence whose identifier is 15365. This T1 control element has the DNA sequence

20

Seq. Id. = 245 Position = 1 to 336

25

AAAATTTCCGGCAAATCGGCAATTTGCCAAAAATTGAAATTTCCGGCAAATCGGCAA
TTTGTCAAAAATGAAAATTTCCGGCAAATCGGCAAATTGCCGAAAATGAAAATTTCC
GGCAAATCGGCAAACCTCCGGAAGTAAAATTTCCGGCAAATCGGCAATTTGCCATA
AATGAACATTTCCGG...GGCGAAAATTAAAATTTCCGCCATATCGGCAATTTGCCA
AAAAATTAAAATTTCCGGCAAATCGGCAAATTGCCGGAATTCAAAATTTCCGGCAA
CCGGCAAATTGCCGGAAGTCAAAATTTCCGGCAAATCAGCAAATTTGCCGGAATT

30

This double stranded DNA loop is bounded on the right by a T2 control element whose identifier is 15627. This T2 control element has the DNA sequence

Seq. Id. = 246 Position = 1 to 68

TGGCAAACCGGCAAATTGCCGGAATTGAACATTTCCGGCAAATCGGCAATTTGCCGG
5 AATTGAAATTT

There are no genes controlled by this T1/T2 loop.

This long T1/T2 double stranded DNA loop modulates the
10 expression of the following C1/C2 short loops

A C1/C2 short loop on chromosome 3 whose identifier is
15366 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop has the
15 DNA sequence

Seq. Id. = 247 Position = 1 to 60

TGCCGATTTGCCGGAATTTTCATTTTCGGCAATTTGCCGATTTGCCGGAATTTTC
20 ATT

This T1-T2 loop also modulates the C1/C2 short loops
numbered 15366 to 15624

25 A C1/C2 short loop on chromosome 3 whose identifier is
15625 controls the expression of the genes of one or more
other T1/T2 long loops. This C1/C2 short loop has the
DNA sequence

30 Seq. Id. = 248 Position = 1 to 54

TCAAGCAAATTGTCAAATTCGCGGAATAAACATTTCCGGCAAATCGGCAAATT

The expression of genes in this T1/T2 long loop is controlled by the following C1/C2 short loops.

5 A C1/C2 short loop on chromosome 4 whose identifier is 16760 controls the expression of the genes in this T1/T2 long loop. This C1/C2 short loop is expressed as a RNA single strand that is 3'UTR to the gene T23E1.2 and has the DNA sequence

10 Seq. Id. = 249 Position = 1 to 124

GGCAAATTGCCGAAATTGAACATTTCCGGCAAATCGGCAAATTGCCGGAATTGAACA
TTTCCGGCAAATCGGCAAATTGCCGGAATTGAACATTTCCGGCAAATCGGCAAATTG
CCGGAATTGA

15

The match between the T1 sequence and the C1/C2 sequence is

Seq. Id. = 249 Position = 22 to 52

20

ATTTCCGGCAAATCGGCAAATTGCCGGAATT

The match between the T2 sequence and the C1/C2 sequence is

25

Seq. Id. = 249 Position = 35 to 75

CGGCAAATTGCCGGAATTGAACATTTCCGGCAAATCGGCAA

30

Abstract

An algorithm has been developed to identify four DNA sequences of 20 bases or more that form a structure called a connectron. Two sequences C1 and C2 are adjacent to each other. These sequences are expressed as RNA in the 3'UTR of some genes in many prokaryotic, archaea and eukaryotic genomes. The other half of a connectron is two DNA sequences T1 and T2 that are on the same chromosome and range in distance from each other by about 1kb to 105kb. The C1 sequence is identical to the T1 sequence and the C2 sequence is identical to the T2 sequence. C1/C2 and T1-T2 can be on different chromosomes. The C1/C2 RNA sequence of the gene transcript finds the two double-stranded DNA sequences T1 and T2. The single-stranded RNA and double-stranded DNA then form a triple-stranded Hoogsteen helix of the RNA/DNA/DNA variety. Because the C1 sequence is adjacent to the C2 sequence, the T1 sequence is made spatially adjacent to the T2 sequence in a compact X-shaped structure. Chromatin particles form as compact 30nm assemblies in the DNA between T1 and T2 thus eliminating the intervening genes from promotion and expression. Connectrons remove sets of genes from expression and thus modulate the behavior of many types of cells.

References

1. "Functional and structural units in the chromomere",
Resnik, N. A.; Yampol, G. P.; Kiseleva, E. V.;
Khristolyubova N. B. and Gruzdev, A. D., Genetica 83:293-
5 299, 1991.
2. "A Silence that speaks volumes", Gura, T., Nature
404:804-808.
3. "RNA Interference", Sharp, P. A. and Zamore, P. D.,
Science 287:2431-2433.
- 10 4. "Genetic Requirements for Inheritance of RNAi in C.
elegans", Grishok, A.; Tabara, H.; Mello, C. C., Science
287:2494-2496.
5. "An RNA-directed nuclease mediates post-
transcriptional gene silencing in Drosophila cells",
15 Hammond, S. M.; Bernstein, E.; Beach, D. and Hannon, G.
J., Nature 404:293-296.
6. "A genetic link between co-suppression and RNA
Interference in C. elegans", Ketting, R. F. and Plasterk,
R. H., Nature 404:296-298